THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 67, PART 1 | 1 January 1954 | No. 409 B

EDITORIAL

The total number of communications accepted and published in parts A and B of the *Proceedings* amounted to 370, this number being made up of 243 Papers, 35 Research Notes and 76 Letters to the Editor. 15% of the manuscripts submitted for publication were rejected on the advice of referees. To give the average time required for publication of material can be misleading, since as much as a few months' delay may occur in cases where an author is carrying out an extensive revision which he feels the referees' reports require. A better indication of the speed of publication is the three to four months which is required by a well prepared manuscript which meets with no serious criticisms of content from referees.

As previously, Papers have normally been sent to two referees, and Research Notes and Letters to one, additional referees being consulted only in cases where a dispute or doubt arises.

The policy of the Papers Committee with regard to Letters to the Editor is that announcements of a preliminary nature prior to full publication will not normally be accepted for the *Proceedings*, and referees are asked not to recommend publication of such communications unless a strong case exists.

A recent Notice to Members, which was circulated to Fellows of the Society, was worded in a manner which could easily be misinterpreted. To avoid any doubt which may exist, the Papers Committee desires it to be repeated that both Fellows and Non-Fellows of the Society may submit their own manuscripts for publication. There is no obligation for any author to have a manuscript communicated by a Fellow of the Society. When an article is communicated by a Fellow of the Society, it is of assistance to the Papers Committee and referees to know whether the communicator of the paper personally knows the work described and whether he also recommends it for publication.

H. H. HOPKINS,
Honorary Papers Secretary.

Further Measurements on the Optical Properties of Lead Sulphide, Selenide and Telluride

By D. G. AVERY
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 4th August 1953

Abstract. Earlier work on the optical properties of PbS, PbSe and PbTe is briefly reviewed. Measurements, using reflection techniques, of the optical constants of all three compounds are reported for wavelengths up to 6 μ. The variation of the optical constants with temperature for some specimens of PbS is described and discussed, and the effect on the constants of baking the specimens in O_2 at a few millimetres pressure is described.

§ 1. Introduction

Earlier work on crystals of lead sulphide, selenide and telluride, including transmission experiments (Paul, Jones and Jones 1951, Gibson 1952, Paul and Jones 1953) and reflection methods (Avery 1951, 1953), has revealed the main features of their optical properties. The absorption spectra were found to consist of a strong absorption band with a peak in the ultra-violet or visible regions and with a long-wavelength tail to this band extending to a sharp edge at a wavelength corresponding to the long-wavelength limit of photoconductivity in the material.

Beyond this edge, where the absorption coefficient falls to 100 cm^{-1} or less, a further slow increase in the absorption is attributable to absorption by the free electrons in the crystals. The behaviour of the refractive index, apart from one minor inconsistency, is roughly in accord with simple classical dispersion theory (e.g. Wood 1933).

In work on a variety of specimens of PbS the optical constants have been found to be independent of the concentration of impurities in the samples over a range of concentrations of 200 : 1; the same independence holds for PbSe and PbTe, though the limited number of samples available has not included such a wide range of impurity concentrations. These results all combine to suggest that the observed optical properties are truly characteristic of the material, rather than any impurity or defect phenomenon. A theoretical treatment of the electronic band structure of PbS, leading to results in fair agreement with these experiments, has been given recently by Bell *et al.* (1953).

The experiments reported here include the extension of the reflection techniques, earlier used by the author, to wavelengths up to 6 μ; this wavelength range includes the long-wavelength edges of all three compounds at room temperature. In addition, measurements on the effects of temperature and oxidation on the optical constants of PbS are reported. The results confirm the intrinsic nature of the observed optical properties and, in particular, enable a fairly complete picture of the optical behaviour of PbS to be drawn.

§ 2. Experimental Techniques

The optical methods used have already been described (Avery 1952) and will not be discussed further, except to draw attention again to the limitations of the method for low values of the absorption. If the absorption is expressed as the imaginary part $2n^2k$ of the dielectric constant $\{n(1-ik)\}^2$, then when k is small serious errors can be introduced by non-parallelism of the incident light and non-flatness of the specimen surfaces. Since the earlier measurements were reported an attempt has been made to reduce these errors by the application of a correction factor dependent on the effective variation in incidence angle. By this means improved agreement between the results of reflection and transmission experiments at low absorptions has been obtained. Nevertheless, with a normal crystal surface for values of $2n^2k=1$, the errors may be as great as 100%, falling to $\pm 10\%$ for $2n^2k=4$ and $\pm 5\%$ for $2n^2k=10$ ($n \simeq 4$). The errors in refractive index measurement are generally less than 5%.

Two minor modifications to the apparatus have been made. The PbS cell used as detector has been replaced by a Hilger Schwarz vacuum thermopile for the longer wavelengths, and the temperature of the specimen can be varied between 15°C and 350°C by a small heating coil on the specimen holder. The temperature was measured with a thermocouple embedded in the specimen holder. To avoid oxidation of the crystals at the higher temperatures the apparatus was placed in a vacuum jacket and measurements were made at pressures of about 2×10^{-5} mm Hg. It was found that heating to 300°C at this pressure for one hour produced no observable permanent changes in the room temperature optical constants, and so it was assumed that serious oxidation did not take place. The specimen could also, if required, be baked in oxygen at these temperatures, the oxygen pressure being roughly measured with a glow discharge.

Cleavage faces of crystals only have been examined, polished surfaces having been shown to give unreliable data. The majority of the crystals were synthetic, and were grown in this laboratory by W. D. Lawson using techniques described elsewhere (Lawson 1951, 1952). In addition some natural crystals of galena were available.

The measurement of the temperature variation of the optical constants involved the measurement of changes in small values of $2n^2k$, in the presence of the possibly large errors referred to above. To reduce these errors to a minimum the specimen surfaces should be as large and as flat as possible, and the only suitable samples were found from specimens of natural galena and one synthetic single crystal of PbS. This part of the work was for this reason restricted to PbS.

§ 3. Room Temperature Measurements on PbS, PbSe and PbTe

In figure 1 the absorption spectra of these compounds are shown, the ordinates being values of the imaginary parts of the dielectric constant. The long-wavelength absorption limits agree within the resolution employed, $\pm 0{\cdot}1\,\mu$, with the results of the other workers quoted above, and the wavelengths of peak absorption agree with measurements on the absorption of thin films of these materials by Gibson (1950). The graphs shown are typical of the materials, the variation from specimen to specimen being no more than experimental error.

The refractive indices of the three compounds at 6 μ have the values PbS $4{\cdot}19 \pm 0{\cdot}06$, PbSe $4{\cdot}54 \pm 0{\cdot}10$ and PbTe $5{\cdot}48 \pm 0{\cdot}10$. It is not likely that further dispersion exists at longer wavelengths due to these absorption bands.

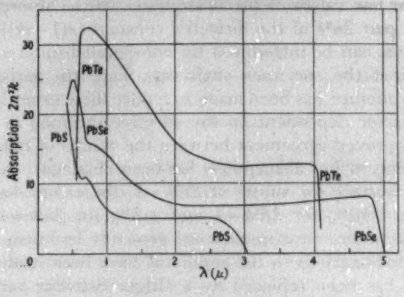


Figure 1. Absorption as function of wavelength for PbS, PbSe and PbTe.

The slight difference observed (Avery 1953) between the dispersion of p- and n-type samples of PbS has been confirmed, and appears to exist in all three compounds. A tentative explanation will be offered below.

§ 4. The Optical Constants of PbS

4.1. *Temperature Effects*

The variation with temperature of the absorption of a PbS crystal is shown in figure 2 (*a*). In this diagram the imaginary part $2n^2k$ of the dielectric constant is plotted against wavelength for various temperatures in the range 15–312°c. The immediately obvious feature is that the characteristic long-wavelength tail progressively vanishes as the temperature increases. Within the error limitations stressed in § 2 above, these variations are repeatable from crystal to crystal. Figure 2 (*b*) shows the relation between an absorption 'edge' (defined as the energy of an incident photon for which $2n^2k$ has fallen to half its room temperature plateau value) and temperature for four different crystals. Within the rather limited purity range from which suitable specimens have been found, this graph shows that the effects are intrinsic to the materials and independent of the impurity concentration.

The rate of change of absorption edge position with temperature between 15° and 125°c is found to be 4×10^{-4} ev/°c, in fair agreement with the values quoted by Gibson (1952) and Paul and Jones (1953), and with a value found for the temperature shift of the long-wavelength limit of photoconductivity by Moss (1949). At temperatures higher than this, however, the rate of change of edge position increases, as opposed to the opposite effect reported by Gibson.

Figure 3 shows the variation of the absorption coefficient K (cm^{-1}) with incident photon energy in ev for various temperatures. The values of K are calculated from these measurements of n and k; also included are some values obtained from direct transmission measurements by Gibson (1952) together with some recent data on thin film absorption by Vernier (1953). For the purpose of this comparison Gibson's curves have all been moved 0·1 μ towards

longer wavelengths, to bring the room temperature curves into close agreement. Since the resolution of Gibson's monochromator was $\mp 0{\cdot}2\,\mu$, as opposed to $\pm 0{\cdot}1\,\mu$ in the present case, this does not seriously misrepresent the former's results. Figure 3 shows clearly the magnitude of the room temperature absorption edge at 0·4 ev, and also the regular movement of the edge towards shorter wavelengths as the temperature increases to 100°c. At higher temperatures this edge becomes progressively less vertical and the shape of the absorption band changes, a possible form for the band at 330°c being indicated. It is this change in shape of the absorption edge that causes the difference between the

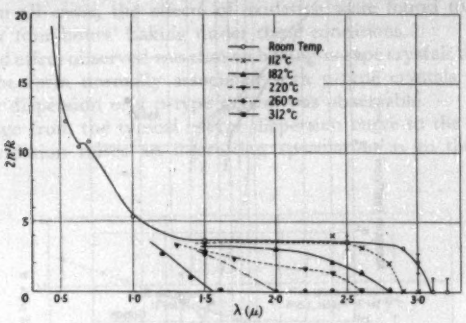


Figure 2 (*a*). $2n^2k$ as function of wavelength for various temperatures for PbS.

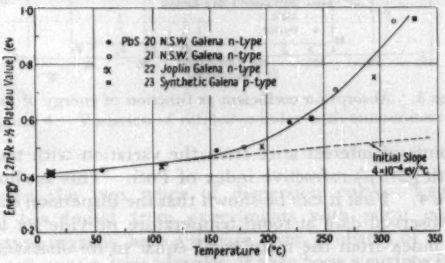


Figure 2 (*b*). Energy at half absorption as function of temperature.

high temperature values for the rate of shift of the edge obtained by Gibson and by the author, the former being measured at a low absorption level ($K \simeq 80$ cm^{-1}) and the latter at a much higher level ($K \simeq 10^4$ cm^{-1}).

Study of the absorption curves of either figure 2 (*a*) or figure 3 suggests that as a rough approximation the process can be split into two separate bands, the less intense band in the $2\,\mu$ region disappearing as the temperature is raised. Dispersion theory calculations based on this two-band approximation, and taking into account the corrections to the $2n^2k$ values mentioned in § 2, lead to values of the oscillator strengths for the two bands of 3·0 ($\lambda_{max} = 0{\cdot}37\,\mu$) and 0·014 ($\lambda_{max} = 2\mu$).

It might be suggested that the disappearance of the subsidiary band as the temperature increases is due to thermal excitation of the electrons lying in the levels associated with the band at room temperature. However, even at 300°c the fraction of the electrons in the full band, lying at levels about 0·5 ev below the conduction band, which are likely to be thermally excited is still very small (only e^{-5} of the total number), and the loss of this small number will not account for the larger change in absorption coefficient observed.

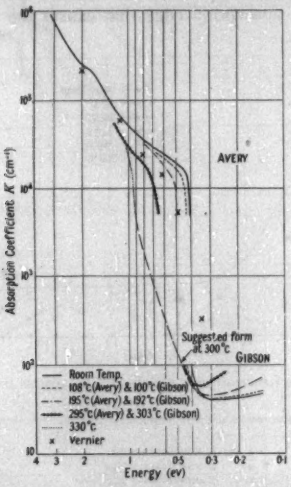


Figure 3. Absorption coefficient as function of energy of radiation.

Further points of interest arise from the variation with temperature of the long-wavelength ($>3\,\mu$) refractive index of PbS. This is shown for a typical crystal in figure 4. First it can be shown that the dispersion theory calculations which fit the observed data at room temperature provide for a contribution to this refractive index from the main band equal to its observed value at 300°c when the subsidiary band has disappeared (the main band can be considered relatively unchanged by this comparatively small temperature rise). Secondly, we can use the theoretical relationship between refractive index and the spectral limit of photoconductivity derived by Moss (1951) to estimate the value of the rate of change of refractive index with temperature. From Moss's equation $n^4/\lambda_c = \text{const}$, if we equate photoconductive threshold with absorption edge, we find $dn/dT = -10^{-3}/°\text{c}$. This compares with the experimentally observed value of $-(5{\cdot}5 \pm 1{\cdot}5) \times 10^{-4}/°\text{c}$.

4.2. *Oxidation Effects*

In view of the importance of oxygen treatment in the manufacture of PbS photoconductive cells (e.g. Smith 1953), observations were made of the effect of baking crystals of PbS in oxygen at a pressure of about 1–5 mm Hg at a

temperature of 200°c. It was found that two principal irreversible changes in the optical constants took place. Firstly a secondary absorption band appeared, with its peak at a wavelength between 1·0 and 1·3 μ. This effect was only reproducible from crystal to crystal in so far as the wavelength of the maximum added absorption always occurred in the 1·0–1·3 μ region. The magnitude of the increase and the rate at which it appeared varied considerably between the three crystals examined. In one case the value of $2n^2k$ at 1·2 μ had increased by 4·6 after four hours, whereas for the other two crystals the corresponding values were 2·0 and 1·3. It is quite possible that this variation is due to the lack of accurate control or measurement of the oxygen pressure employed. In all cases, the effects of oxidation were found to be practically complete after four hours' baking under these conditions.

The second effect observed was that on baking n-type crystals their dispersion changed to the form normally associated with p-type crystals (see § 3). No change in the dispersion of a p-type crystal was observable.

This change from the typical n-type dispersion curve to the p-type characteristic on oxidation raises an interesting speculation as to the cause of the

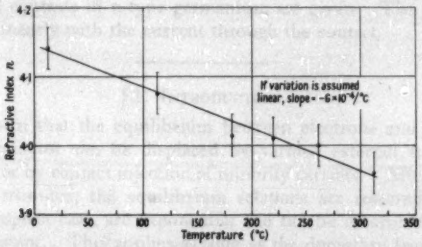


Figure 4. Variation of refractive index with temperature.

difference between the two types of dispersion curve. Lawrance (private communication) has found from experiments with point contacts on PbS surfaces that p-type crystals have a surface layer, probably of oxidized material, about 1 μ thick, whereas n-type samples, if they have a surface layer at all, have a very much thinner layer. Since the penetration distance of the incident radiation is about 1 μ, it may be that this oxidized layer on p-type samples affects the reflection properties sufficiently to change the form of the dispersion curve, whilst the change in the absorption spectrum is still too small to be observed in the presence of the much greater errors in the absorption measurements.

Acknowledgments

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller, Her Majesty's Stationery Office, for permission to publish this paper.

References

AVERY, D. G., 1951, *Proc. Phys. Soc.* B, **64,** 1087; 1952, *Ibid.*, **65,** 425; 1953, *Ibid.*, **66,** 133.
BELL, D. G., HUM, D. M., PINCHERLE, L., SCIAMA, D. W., and WOODWARD, P. M., 1953, *Proc. Roy. Soc.* A, **217,** 71.
GIBSON, A. F., 1950, *Proc. Phys. Soc.* B, **63,** 756; 1952, *Ibid.*, **65,** 378.
LAWSON, W. D., 1951, *J. Appl. Phys.*, **22,** 1444; 1952, *Ibid.*, **23,** 495.
MOSS, T. S., 1949, *Proc. Phys. Soc.* B, **62,** 741; 1951, *Proc. Phys. Soc.* A, **64,** 590.
PAUL, W., and JONES, R. V., 1953, *Proc. Phys. Soc.* B, **66,** 194.
PAUL, W., JONES, D. A., and JONES, R. V., 1951, *Proc. Phys. Soc.* B, **64,** 528.
SMITH, R. A., 1953, *Advances in Physics* (*Phil. Mag.* Suppl.), **2,** 321.
VERNIER, P., 1953, *J. Phys. Radium*, **14,** 175.
WOOD, R. W., 1933, *Physical Optics*, 3rd Edn (London : Macmillan).

Measurement of Minority Carrier Lifetime and Contact Injection Ratio on Transistor Materials

By A. MANY*

Department of Physics, University of Reading

Communicated by R. W. Ditchburn; MS. received 1st September 1953

Abstract. A method of measuring minority carrier lifetime and injection ratios of contacts on transistor materials is described. The lifetime is obtained by observing the decay of resistance of a filament throughout the duration of an injecting pulse. The use of a bridge circuit, incorporating a resistance–capacitance network which is electrically analogous to the filament, enables the lifetime to be read off a calibrated dial. The range of lifetime, which can be measured with the equipment in its present form, extends down to 1 μsec, the accuracy being usually about 5%. The method is also capable of determining minority carrier mobilities. Examples of measurements of lifetime and injection ratio of soldered contacts in n-type germanium are given. The injection ratio is found to rise linearly with the current through the contact.

§ 1. Introduction

It is well known that the equilibrium between electrons and positive holes in a semiconductor can be displaced by various external stimuli, e.g. by illumination or by contact injection of minority carriers. After the cessation of the external stimulus, the equilibrium relations are restored by processes which, in the simplest case, are exponential and can be described in terms of a single decay constant. This applies as long as the departure from equilibrium is small. The decay is then independent of the magnitude of the original disturbance. The effective lifetime so determined is made up of two contributions which are due to bulk and surface recombination respectively. The relative importance of these contributions depends on the surface condition and the size and shape of the specimen.

Measurements of lifetime are of great importance in connection with the study of lattice disorder and surface conditions. They are also essential in the assessment of transistor materials. A number of methods have been described which can be used for this purpose (Haynes and Shockley 1951, Goucher 1951, Navon, Bray and Fan 1952). The present method is similar, in some respects, to that described by Navon, Bray and Fan (1952). These authors have evaluated the carrier lifetime from observations of resistivity, immediately following the cessation of an injecting voltage pulse. In the present method the resistance of the specimen is monitored continuously throughout the duration of the injecting pulse, and a bridge circuit is employed which incorporates a resistance–capacitance network electrically analogous to the specimen. This dispenses with a good deal of elaborate analysis and enables the carrier lifetime to be read off a calibrated dial. The method is capable of high accuracy and is suitable

* Now at the Hebrew University, Jerusalem.

for the measurement of lifetimes which are outside the useful range of the methods hitherto employed. It also leads to a measurement of injection ratio and carrier mobility. The experiments described have been carried out on n-type germanium. The method should be equally applicable to p-type specimens and to other transistor materials.

§ 2. Theory

The quantitative considerations will be limited to specimens of simple geometry, e.g. filaments of rectangular cross section A and length L. The treatment is given in terms of n-type material, in which n_0 and p_0 represent the equilibrium concentrations of electrons and holes respectively. Thus $p_0 \ll n_0$. Departures from equilibrium are assumed to decay exponentially with a time constant τ, the effective lifetime of the filament. The conductance G_0 of such a filament is given by

$$G_0 = \boldsymbol{e}(\mu_n n_0 + \mu_p p_0)(A/L) \qquad \ldots\ldots(1)$$

where μ_n and μ_p are the electron and hole mobilities. It has been shown (Shockley 1950) that the effect of additional minority carriers on the conductance is independent of the manner in which these additional carriers are distributed over the specimen, as long as the local concentration p satisfies $p \ll n_0$, and as long as diffusion effects can be neglected. Thus, if neutrality prevails, the increase of conductance ΔG of the specimen is given by

$$\Delta G = \boldsymbol{e}(\mu_n + \mu_p) L^{-2} P_{tot} \qquad \ldots\ldots(2)$$

where $P_{tot} = \int p\, dx\, dy\, dz$ is the total number of added holes. Combining eqns (1) and (2) we get

$$\frac{\Delta G}{G_0} = \frac{(b+1)P_{tot}}{u(bn_0 + p_0)} \qquad \ldots\ldots(3)$$

where u is the volume of the filament and b the ratio of electron to hole mobility.

In the present case the added hole concentration is introduced by injection through one of the end contacts. For the purposes of the experiment this contact must have a low injection ratio γ so as to fulfil the condition $p \ll n_0$ at all points of the specimen. The second contact, which is inevitably operated in the reverse direction, should be non-rectifying and must, even under these conditions, have a low contact resistance, so that eqn (1) may be satisfied. This condition is also necessary from other considerations, which will be discussed below. The injection takes place during a square pulse of constant current I. Thus $\gamma I/\boldsymbol{e}$ additional holes are injected in unit time and are swept along the filament by the applied field. At the same time the holes tend to recombine and those which are still free after traversing the filament in a transit time τ_t are removed through the end contact. The transit time depends on the applied field, and in view of the condition $p \ll n_0$ it can be assumed that there is no appreciable disturbance of the field which might alter with time.

Let the onset of the current pulse coincide with $t=0$. At any instant t' the number of holes injected in an interval dt' is $(\gamma I/\boldsymbol{e})dt'$. At any later time t, such that $t' < t \leqslant \tau_t$, only those minority carriers which have not yet recombined are able to affect the conductance of the filament. The number of such uncombined holes is given by $(\gamma I/\boldsymbol{e})dt' \exp[-(t-t')/\tau]$. The total number of additional holes at time t is thus obtained by integrating the last expression from $t'=0$ to $t'=t$. This leads to

$$P_{tot} = (\gamma I\tau/\boldsymbol{e})[1-\exp(-t/\tau)], \qquad \text{for } t \leqslant \tau_t. \qquad \ldots\ldots(4)$$

In the case of $t > \tau_t$ the term t in the exponential of the last equation should be replaced by τ_t. It will be seen that, if $\tau_t \gg \tau$, P_{tot} rises exponentially throughout the duration of the pulse and tends to the limiting value $\gamma I\tau/\mathbf{e}$. If, on the other hand, τ_t is comparable with or shorter than τ, P_{tot} first rises exponentially until $t = \tau_t$ and then becomes constant at $P_{tot}(\tau_t)$. The rate of carrier injection is then balanced by the recombination rate and the rate of removal at the end-contact.

It is now possible to substitute for P_{tot} in eqn (3). This gives the variation of conductance as a function of time:

$$\frac{\Delta G(t)}{G_0} = \frac{(b+1)\gamma I\tau}{\mathbf{e}(bn_0+p_0)u}[1-\exp(-t/\tau)], \qquad \text{for } t \leqslant \tau_t. \quad \ldots\ldots(5)$$

Thus τ and γ can be evaluated from measurements of ΔG as a function of time.

A low γ value was assumed for eqn (2) to be satisfied. This is also necessary in order to make the concentration gradients of the added carriers, and hence the diffusion currents, negligible. For high γ values the diffusion currents may be comparable with drift current (Banbury 1953 b), and interpretation of P_{tot} in terms of conductivity modulation would no longer be applicable.

§ 3. Experimental Method

The condition $p \ll n_0$ leads to $P_{tot} \ll un_0$ and hence to $\Delta G/G_0 \ll 1$ (eqn (3)). In terms of filament resistance this means also $\Delta G/G_0 = -\Delta R/R_0$, and, since a constant current source is employed, the voltage V across the filament is proportional to its resistance at any time. We thus have

$$\Delta G(t)/G_0 = -\Delta V(t)/V_0 \qquad \ldots\ldots(6)$$

where V_0 is the voltage across the specimen at instant $t = 0$. Since $\Delta V(t)/V_0$ is small, it is impracticable to observe $\Delta V(t)$ by means of an amplifier connected directly across the specimen. In this case V_0 would be similarly amplified and would give rise to overloading. These particular difficulties can be avoided by the use of a simple bridge method. (The author is indebted to Mr. R. E. Burgess of the Radio Research Station for drawing his attention to this principle.)

For the special investigations here under discussion the circuit shown in figure 1 was devised. The filament constitutes one arm of the bridge, the others being made up of ordinary linear resistors. The resistance R_1 in series with the filament is made large compared with R_0 so that the constant voltage generator used for the experiment becomes effectively a constant current source. By proper adjustment of the resistance r_1, the bridge between points a, b is balanced for the *initial* resistance R_0 of the filament, i.e. the resistance corresponding to $t = 0$, just before injection becomes effective. The exponential change of filament conductance with time shows itself later as an unbalanced signal between points a and b of the bridge. This signal can be amplified and displayed on a cathode-ray oscilloscope as a function of time. Oscillograms so obtained for various applied currents through the filament are shown in figure 2. Analysis of such curves would give the lifetime of the injected carriers. However, this procedure is laborious and depends on the linearity of the amplifiers. An alternative method which is much faster and more accurate will now be described.

The voltage decay across the filament is simulated by the exponential decay of current charging a condenser through a resistance. A series combination of a resistance R and a capacitance C is connected between points b and d of the circuit, and the values are chosen so that $R \gg r_0$. If v_0 is the voltage across

points c and e at $t=0$ and $\Delta v(t)$ the decay of this voltage as a function of time, then, using the terms given in figure 1, we have

$$\Delta v(t)/v_0 = -(\Delta r/r_0)[1-\exp(-t/RC)]. \quad \ldots\ldots(7)$$

v_0 is equal to V_0 as a result of the initial balance of the bridge. The product RC and the ratio $\Delta r/r_0$ are now adjusted until complete balance is obtained across points a and c throughout the duration of the pulse, i.e. for all values of t smaller than τ_t. On the cathode-ray oscilloscope this balance is indicated by the appearance of a straight horizontal line. The lifetime of the minority carriers and the

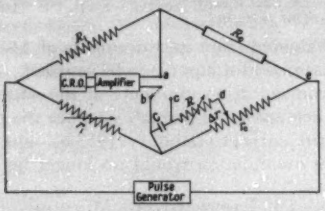


Figure 1. Bridge circuit employed for measuring lifetime and injection ratio.

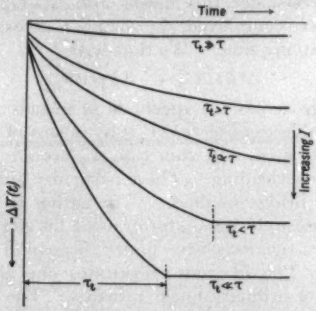


Figure 2. A typical set of voltage decay curves for various currents through the filament. (*Note*. Intercept on ordinate shows slight initial diminution due to properties of the injecting contact; it later remains constant.)

injection ratio can then be evaluated from

$$\tau = RC \quad \ldots\ldots(8)$$

and

$$\frac{\Delta r}{r_0} = \frac{(b+1)\gamma I\tau}{e(bn_0+p_0)u}. \quad \ldots\ldots(9)$$

Values of τ can be directly marked on a dial. It should be noted that the bridge balance is maintained only during the pulse. After the cessation of the pulse, the capacitor C is charged and thus the voltage across points c and e decays while C discharges through R. The voltage across the filament, on the other hand, drops immediately to zero or, more precisely, to a very low value corresponding to the 'floating potential' at the injecting contact. This matter will be further discussed below. Thus, the straight horizontal line which corresponds

to balance on the cathode-ray oscilloscope is followed by an exponential decay after the cessation of the injecting pulse. This decay is of no importance for the experiment, but its existence implies that the use of a cathode-ray oscilloscope display is essential and that no simpler null indicator could be employed.

§ 4. Results

4.1. *Lifetime*

Measurements were carried out on several germanium samples, with various surface treatments, giving a range of values from a few microseconds to several hundred microseconds. The lowest lifetime measured on the present instrument is about 1 μsec. This limitation arises from the particular circuit assembly employed, and it is believed that with more careful design the limit could be lowered to 0·1 μsec or less. The resistance decay was usually found to be an exponential with a single time constant. In such cases the setting of the resistance–capacitance network for balance could be made to an accuracy of within about 5%, which compares favourably with other published methods.

As mentioned above, the excess carriers recombine both in the bulk and at the surface of the filament. The contribution of each process to the effective carrier lifetime has been worked out for a rectangular specimen (Shockley 1950). This analysis can readily be extended to cover the present case. The decay is expected to be composed of a predominant exponential term and a number of other terms having much shorter time constants. The nature and composition of these transients should depend on the initial conditions at $t=0$, i.e. on the distribution of the injected carrier current density over the contact area. This initial distribution adjusts itself rapidly to that imposed by the geometry and surface condition of the filament. These transients have been observed experimentally in some cases: the slow decay was preceded by a short initial decay of time constant many times smaller.

The predominant decay constant corresponds to the effective filament lifetime, and is given by (Shockley 1950),

$$1/\tau = 1/\tau_b + 1/\tau_s \qquad \ldots\ldots(10)$$

where τ_b is the bulk lifetime in the absence of surface recombination and τ_s is the surface lifetime which can be calculated in terms of the surface recombination velocity s and the dimensions of the filament; τ_s increases by reducing the former and increasing the latter. Thus for measuring bulk lifetime a large filament with well-etched surfaces should be used. The surface recombination velocity, on the other hand. would most readily be evaluated by using small section filaments, such that $\tau_s \ll \tau_b$, so that effectively all recombination takes place at the surface.

In the table are listed some of the measurements taken on various etched filaments of n-type germanium material. The first row corresponds to pulse

Lifetime Measurements (in microseconds)

Pulse method, with soldered contacts	40	65	105	210
By diffusion length	30	40	80	170
Pulse method, with point contacts	45	75	100	170

measurements using soldered end contacts. The second row gives estimated lifetimes obtained on the same filaments from measurement of the diffusion range of optically generated carriers using a conventional probe method. The

latter values are only approximate; they are always lower than those obtained by the pulse method. This is due to the large and indeterminate contribution of the surface recombination in the diffusion range experiments. The third row corresponds to pulse measurements using point contacts, and will be discussed below.

Measurements were also carried out on filaments of long lifetime germanium with roughened surfaces. These surfaces have s values above 10 000 cm sec^{-1}. For sufficiently small cross sections the effective filament lifetime should be then equal to the surface lifetime. Moreover, the latter is expected to have only slight dependence on s, being mainly determined by the dimensions of the filament cross section. A filament with an initial cross section of 2 mm × 2 mm was used. One dimension was gradually reduced, by grinding, down to 0·4 mm. Lifetimes were determined for several thicknesses, in each case both for a ground

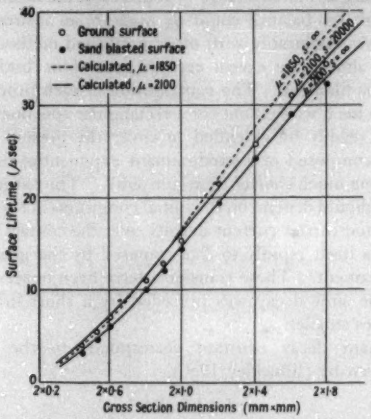


Figure 3. Measured and calculated lifetimes for filaments of varying cross section.

surface (using Aloxite powder No. 50) and for a sand-blasted surface (using Carborundum No. 240). The results are shown by the points in figure 3. The broken curve represents the corresponding lifetimes calculated for $s=\infty$ (Shockley 1950), and using the most recent value for the hole drift mobility, i.e. 1850 cm^{2}v^{-1}sec^{-1} (Conwell, private communication). It is seen that the measured results are consistently lower than the calculated values. It was found that if the value 2100 cm^{2}v^{-1}sec^{-1} were used for drift mobility, a much better fit would be obtained. This is shown by the full lines in figure 3, which represent calculated lifetimes corresponding to $s=20\,000$ cm sec^{-1} and $s=\infty$ respectively. It should be noted that the values of $\mu=2100$ and $s=20\,000$ for a ground surface, implied by this fit, are in agreement with results quoted by Navon, Bray and Fan (1952).

The method was also used for the evaluation of surface recombination velocity after various surface treatments. It was found, for example, that very well etched surfaces deteriorate with time when exposed to atmospheric conditions.

In one particular case s increased from 200 cm sec^{-1} immediately following the application of an etch to 1000 cm sec^{-1} after a few days.

4.2. *Mobility*

The mobility of the injected carriers is given by $\mu = L/E\tau_t$, where τ_t is the transit time through the filament under the field E. As pointed out above, this relation holds only for low injected carrier densities, a condition which can readily be satisfied in the present experiment. τ_t is measured by determining the instant in which the voltage exponential decay, as observed on the cathode-ray oscilloscope, flattens out into a straight horizontal line (see figure 2). In this way hole mobilities were measured in a number of n-type germanium filaments. The results agree, within the accuracy of measurement, at present only about 20%, with the accepted values.

4.3. *Injection Ratio*

γ is determined by measuring $\Delta r/r_0$ and employing eqn (9). Values of γ for different soldered contacts were determined as a function of total current I through the contacts. Some of these results are shown in figure 4. It is seen

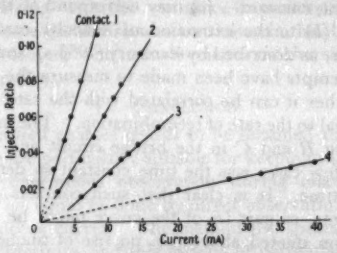


Figure 4. Injection ratios of four soldered contacts as a function of current through the filaments.

that γ increases linearly with current. This behaviour was found for all soldered contacts examined. No satisfactory theory of poorly injecting contacts is so far available for the interpretation of these results.

It was assumed in the analysis leading to eqn (4) that the injection ratio is constant throughout the duration of the pulse. However, it could be conceived that γ might depend on the injected carrier concentration near the contact barrier, and might thus vary with time. The validity of the original assumption was therefore checked by repeating the measurement of γ while the area near the contact was illuminated to increase the carrier concentration in that region. The light intensity was chosen so that the optically generated carriers reduced the filament resistance by at least the same amount as the injected carriers under comparable conditions. During this experiment γ was found to be unaltered.

4.4. *End Contacts*

As mentioned above, it is necessary that the injecting contact should have a low γ, and that the other contact be non-rectifying. These requirements can be fulfilled by using soldered contacts. As was shown above (figure 4),

γ rises with applied current, and therefore almost any injecting contact is satisfactory, as the desired range of γ values can usually be obtained by proper choice of the current. The non-rectifying contact is obtained by careful soldering; in some cases it was found necessary to nickel-plate before soldering. The suitability of the contact is checked on the instrument itself by ensuring that the filament resistance at time $t=0$ (i.e. before injection) does not rise with applied voltage. For non-rectifying contacts the total resistance of the specimen at zero time should be entirely independent of applied voltage. In fact the injecting contact is slightly rectifying and its resistance will diminish somewhat with increasing forward voltage. A typical set of decay curves is as shown in figure 2.

If the resistance of the other end contact is not entirely negligible, as it should be for purposes of the experiment, this fact can be recognized by an observed increase of total resistance at zero time with increasing applied voltage. Such an increase corresponds to the general behaviour of rectifying contacts biased in the reverse direction. The curves observed on the cathode-ray oscilloscope are then very complex. They also vary considerably from contact to contact and over short periods of time. The curves are usually composed of three parts: (*a*) a fast rise in resistance with time, (*b*) a slower rise, and (*c*) a decay of resistance with a very long time constant. (*a*) may correspond to the charging up of the barrier capacitance, (*b*) to the extraction of minority carriers from the region adjoining the barrier, as described by Banbury (1953 a), and (*c*) to thermal effects in the barrier. Attempts have been made to measure the time constant of part (*b*) to find out whether it can be correlated with the rate of carrier generation, which should be equal to the rate of recombination. The measurement is carried out by interchanging R and C in the bridge circuit and balancing part (*b*) as before. In only a few cases was the time constant so determined found equal to the filament lifetime. It is clear that such contacts must be avoided for lifetime measurements, as part (*c*) of the curve might be mistaken for lifetime decay. The criterion quoted above (i.e. no rise of filament resistance at $t=0$ with applied voltage) is quite reliable for rejecting these contacts.

4.5. *Effect of Pulse Amplitude*

In a few filaments, particularly those having very low s values, the measured effective lifetime was found to increase with applied current. This behaviour was invariably associated with highly injecting contacts, and is not yet properly understood. For such low s values the effective lifetime of the filament is large. In accordance with eqn (4) and the behaviour of γ as discussed above, P_{tot} is expected to increase with increasing current. In the specimens of high γ and long lifetime this increase will be particularly marked and may offer a possible explanation of the dependence of lifetime on current. Such a dependence may result from the presence of centres, perhaps at the surface, in which delayed recombination occurs and whose occupancy is affected by the high values of P_{tot}.

4.6. *Floating Potential*

Due to the presence of the injected carriers near the barrier associated with the injecting contact, a 'floating potential' (Bardeen 1950) exists at this contact. The potential drop across the barrier is in series with, and in the same direction as, the RI drop across the filament. The former is, however, very small compared with the voltage changes produced by injection, and therefore does not interfere with the lifetime measurements. However, when the pulse is switched off,

only the floating potential remains, and a small decaying voltage is observed on the oscilloscope (when the latter is connected between points a and b of the bridge). The decay is at first rapid and approximately exponential. This is followed by a region in which the floating potential falls linearly to a very low value near the limit of observation. The linear part agrees with the logarithmic relation between floating potential and carrier concentration (Bardeen 1950), as the decay of carriers in the filament after the cessation of the pulse is exponential.

4.7. *Measurements on Point Contacts*

The above theoretical treatment does not apply when the injecting electrode is a point contact. Nevertheless an approximate value for the lifetime of the sample can be obtained. The oscillograms obtained by pulsing such a diode consisted of a fast and large amplitude decay lasting for a few microseconds, followed by a relatively low amplitude exponential decay, having a time constant approximately equal to the lifetime of the sample. The fast decay probably corresponds to the introduction of the injected carriers into the spreading resistance region, while the slow decay may be associated with the modulation of the bulk resistance as the carriers move into the bulk. The decay constant is measured by balancing with the resistance–capacitance network the latter part of the curve. Some lifetime results are shown in the third line of the table, and are in rough agreement with measurements on the same filaments, using soldered contacts (top line).

§ 5. Conclusion

The method described enables results to be obtained rapidly and with high accuracy. It is therefore particularly suitable for keeping the surface recombination velocity under continuous observation during changes resulting from surface reactions. This could be done by direct observation on the oscilloscope of the compensated decay curves. The method has special advantages for the determination of low lifetimes, and this may be of particular value for the study of transistor materials other than germanium, for example lead sulphide. The present technique can also be used for the rapid assessment of transistor material in the form of ingots as produced in industry. The use of the diode configuration as described above provides a quick and non-destructive test which may be useful in this connection. In principle the method could be extended to photoconductors which do not necessarily exhibit contact injection. The additional minority carriers would then have to be produced by a synchronized light pulse.

Acknowledgments

Grateful thanks are due to Professor R. W. Ditchburn for placing research facilities at the author's disposal and to Drs. H. K. Henisch and P. C. Banbury for most valuable help and suggestions. The author is also indebted to The Humanitarian Trust and the Friends of the Hebrew University for a grant.

References

Banbury, P. C., 1953 a, *Proc. Phys. Soc.* B, **66**, 50; 1953 b, *Ibid.* B, **66**, 833.
Bardeen, J., 1950, *Bell Syst. Tech. J.*, **29**, 469.
Goucher, F. S., 1951, *Phys. Rev.*, **81**, 475.
Haynes, J. R., and Shockley, W., 1951, *Phys. Rev.*, **81**, 835.
Navon, D., Bray, R., and Fan, H. Y., 1952, *Proc. Inst. Radio Engrs., N.Y.*, **40**, 1342.
Shockley, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).

The Temperature Dependence of the Drift Mobility of Injected Holes in Germanium

By R. LAWRANCE *

Telecommunications Research Establishment, Great Malvern, Worcs.

Abstract. The drift mobility of holes in n-type germanium has been studied as a function of temperature. In the temperature range 100°K to 300°K the mobility varies as $T^{-2\cdot3}$ except when it is dominated by trapping. The effect of trapping on the mobility has been investigated, and a suitable trapping model suggested.

§ 1. Introduction

When a point contact rectifier made from germanium of high purity is biased in the forward (or low resistance) direction, current carriers of the opposite polarity to those normally present in the material may be introduced into the germanium. This process is known as minority carrier injection, and is the basis of all transistor action. Thus, if the germanium be n-type, having electrons as the normal current carriers, positive holes may be injected at the point contact. For most n-type germanium this carrier injection process is extremely efficient, the proportion γ of the current being carried by holes being nominally unity.

While present in the germanium, the minority carriers will recombine with carriers of the opposite sign. This decay process is exponential in time, the time taken for injected carriers to decay to 1/e of their original number being known as the lifetime τ of the carriers. τ may vary between about 0·1 μsec and 1·0 msec in n-type germanium.

The minority carriers may be removed from the germanium by a second point contact rectifier biased in the reverse (or high resistance) direction. This process, known as collection, is fairly efficient, though, due to decay of the carriers, it would be expected that the current at the collector due to the injected carriers would be less than the injection current at the emitter. For holes injected into n-type germanium it is in general found, however, that more than one current carrier passes through the collector for every hole injected at the emitter. Thus, the current gain of the system, as well as the voltage and power gains, may be greater than unity. This is one of the basic differences between the transistor and the thermionic valve.

A current gain $\alpha_c = \partial i_c / \partial i_e$ (V_c constant), where e and c stand for emitter and collector and V_c is the collector voltage, up to a value of 3·1 at room temperature can be explained on the basis of the equation

$$\alpha_c = \gamma\beta(1+b) \qquad \ldots\ldots(1)$$

where β is the hole transfer efficiency and involves τ and the field between the emitter and the collector, and hence $\gamma\beta \leqslant 1$; b is the ratio of the electron mobility

* Now at Electronics Research Laboratories, Department of Supply, Salisbury, South Australia.

to hole mobility, μ_n/μ_p, equal to 2·1 in germanium at room temperature (for general theory see Shockley 1950a).

It is fairly well known that some n-type germanium triodes working under normal conditions have current gain in excess of 3·1, while it has been shown by Sittner (1952) that triodes which exhibit normal current gain under normal conditions may have current gain greater than 3·1 for emitter bias less than about 10^{-5} A. Other workers (E. H. Cooke-Yarborough, private communication) have found current gain of the order of 20 at room temperature and low emitter bias.

Shockley (1950b) put forward two possible mechanisms of high current gain. One was the p–n hook theory, which relied upon the formation of a small p–n barrier under the collector during the process of forming the collector. During forming, a large current pulse is passed through the collector, thus welding it to the germanium surface. The intense heat caused by this pulse was expected to disrupt the crystal lattice in the vicinity of the point, turning it p-type. A small amount of phosphorus was at the same time expected to diffuse from the phosphor-bronze point into the p-type patch, producing a very small n-type patch immediately under the point. Valdes (1952), by the use of a probing method, was able to locate a p-type patch after forming, but was unable to find any evidence of an n-type spot. He assumed that it was too small to observe.

The second of Shockley's mechanisms relied upon distributed traps for holes. Both of the mechanisms were to have the effect of reducing the mobility of holes. Reference to eqn (1) shows that this will have the effect of increasing the mobility ratio b, and hence increasing α.

Sittner (1952) showed that the trap model fitted the high current gain at low emitter currents with tolerable accuracy. He was able to determine a room temperature trap depth of 0·31 ev from his results. Later Hunter *et al.* (1952) showed that both this form of current gain and another which followed the p–n hook model could exist simultaneously in n-type germanium triodes.

If traps for holes exist, as would seem likely from Sittner's remarks, it should be possible to observe the effect of trapping on the drift mobility of holes under suitable conditions of emitter current or temperature. At low temperatures the effect of traps should be more marked than at higher temperatures. Also, the effect of the p–n hook may be eliminated by the use of an unformed collector.

The present experiments have been conducted in order to determine the temperature variation of the drift mobility of holes in n-type germanium and to investigate the effect of trapping and the nature of the traps. The results have been briefly reported elsewhere (Lawrance 1953).

§ 2. Experimental Method

The drift mobility of holes may be measured by the method described by Haynes and Shockley (1951). Basically, the method is as follows. A rod of germanium is cut from an n-type single crystal of suitable resistivtiy, and an emitter and a collector are placed on the surface several millimetres apart. A current is passed down the rod through low resistance end connections in such a direction as to sweep positive holes from the emitter to the collector. The sweeping field may be either direct-current or pulsed to reduce heating at high fields. Holes are introduced into the system by pulsing the emitter.

The method of measurement used in the present experiments was a modification of the Haynes and Shockley method described by Lawrance and Gibson (1952). In this method the emitter is constantly biased in the forward direction with a current of about 0·2 mA, while the sweeping field is supplied as a pulse of up to 50 V amplitude and about 140 μsec in duration. There is sufficient time between pulses for the very small emitter current to build about the emitter an exponential distribution of holes. When the sweeping field switches on, this 'packet' of holes is swept down the rod, reaching the collector after a time t. In practice, the collector signal is displayed on a cathode-ray oscilloscope, the time being measured by superimposing timing marks on the oscilloscope trace. The appearance of the display trace is shown in figure 1.

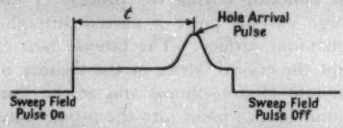


Figure 1. Appearance of the cathode-ray oscilloscope display in the absence of trapping.

For most of the experiments germanium rods about 1 cm long and 0·5 mm × 0·5 mm cross section were cut from single crystal material of 2 ohm cm resistivity and 20 μsec lifetime. In addition, some material of 7 ohm cm, 50 μsec lifetime was used to check the results. These rods were mounted on Perspex blocks without the use of adhesives which may have caused some form of trapping themselves if present on the germanium surface. The method of mounting was to solder one end of a rod to a silver plate screwed to one end of a Perspex block. The other end of the rod was soldered to a phosphor-bronze spring, also screwed to the block. The rod was thus rigid at one end but free to move at the other, so that differential expansion of the Perspex and the germanium would not produce strains in the rod. This is an important consideration, for it was found in early experiments that when both ends of the rod were kept rigid there was a marked effect upon the temperature variation of the mobility of holes. Further mention of this effect will be made.

Low temperature measurements were made with the germanium mounted in a specially constructed micromanipulator designed by Mr. W. H. Mitchell of this establishment. This manipulator, shown in figure 2, contains two carriages to manipulate emitter and collector. It is enclosed in a glass vessel which can be evacuated to prevent frosting at low temperatures. A pair of voltage probes was welded to the germanium a few millimetres apart, the emitter and collector being placed between these probes in order that they should not affect the measured mobility by the formation of p–n hooks. The method of welding voltage probes which are both low resistance and non-rectifying down to liquid air temperatures has been described by Mitchell (1954).

Finally, a thermocouple was soldered to the silver end support of the rod. It was assumed that there was sufficiently good thermal contact between the germanium and thermocouple for the latter to measure the temperature of the former to within a few per cent, whether it be rising or falling. Subsequent measurements confirmed that this was so.

A rod, thus mounted, was placed in a vessel which was evacuated and immersed in liquid air.

§ 3. Some Preliminary Observations

When the vessel containing the germanium was evacuated it was noticed that the hole arrival pulse was reduced in amplitude, often by as much as 100 times. It was thought that this may have been due to the condensation of water vapour on the germanium surface, so a liquid air trap was interposed between the pump and the vessel. This failed to make the hole arrival pulse

Figure 2. Miniaturized micromanipulator.

return to its original height, and it was felt that it was the removal of water vapour from the system rather than condensation which was causing the effect. The following experiments were carried out in order to confirm this view:

(*a*) The vessel was evacuated and air re-admitted slowly through the liquid air trap; the pulse remained small. The experiment was repeated, but this time the air was re-admitted quickly through the trap to reduce its effectiveness; the pulse regained its original height.

(*b*) With dry air at atmospheric pressure in the micromanipulator vessel, and a small hole arrival pulse, the vessel was opened to the atmosphere; the pulse immediately increased in amplitude.

(*c*) The vessel was sealed up at atmospheric pressure without drying the air within it; a large hole arrival pulse was observed. The vessel was then immersed in liquid air and the pulse was observed to diminish. It was assumed that this reduction was caused by the condensation of water vapour from the atmosphere surrounding the germanium on to the cold vessel walls. When the vessel was removed from the liquid air, the pulse slowly regained its original amplitude as the walls warmed up.

(*d*) A drop of water was placed in the vessel which was then evacuated; the pulse remained at its high amplitude until the water froze, after which the pulse immediately fell to a small height. It was assumed that the concentration of water vapour was maintained high until its source was removed when the water froze.

(*e*) A drop of methyl alcohol was placed in the vessel in order to test the effect of a drying agent in the atmosphere surrounding the germanium; the

effect was to reduce the height of the hole arrival pulse, which would not regain its original amplitude even when the alcohol atmosphere was removed. Presumably, water had been removed from the germanium surface and replaced by adsorbed alcohol molecules.

(*f*) It was noticed that the back resistance of the collector was reduced from about 10^6 ohms to about 10^4 ohms when the vessel was evacuated. Thus the reduction in the amplitude of the hole arrival pulse in the above experiments was due to a lowering of the back resistance of the collector when water was removed from the germanium surface. This is tantamount to saying that the surface barrier height was being lowered with removal of water.

Despite the drastic reduction of signal brought about by the evacuation effect, there was still sufficient signal above noise to enable measurements to be made, and all subsequent measurements were made *in vacuo*.

§ 4. Mobility at Low Temperatures

As the temperature of the germanium falls below room temperature the drift mobility of holes increases according to the relationship $\mu \propto T^{-2\cdot3}$. Theoretically the mobility should increase as $T^{-1\cdot5}$ (Shockley 1950a). Dunlap (1950) has measured the temperature variation of the Hall mobility of holes in single crystal p-type germanium, and has obtained a relationship $\mu \propto T^{-n}$, where n lies between 2·0 and 2·3.

At some temperature, dependent upon many factors, the drift mobility of holes ceases to increase with decreasing temperature, and over a range of a few degrees begins to decrease rapidly. This decrease is caused by some sort of trapping agency, and for convenience will be referred to as trapping, though this may not be strictly accurate. In the discussion to follow, this theme will be expanded somewhat.

The onset of trapping is readily observable on the cathode-ray display. Before trapping sets in, the hole arrival pulse is symmetrical about its mean as shown in figure 1. With the onset of trapping, however, this pulse becomes skew as shown in figure 3. This is followed by a bodily increase in the delay

Figure 3. Appearance of the cathode-ray oscilloscope display when trapping begins.

time t. At still lower temperatures the hole arrival pulse becomes extremely skew with a tail extending to transit times much in excess of the transit time of the peak. Finally the collector signal becomes too small to measure, and the mobility as measured from the peak of the hole arrival pulse has by this time fallen by about an order.

The skewness of the hole arrival pulse is thought to be due to the existence of a large range of transit times. An explanation on the basis of trapping theory is immediately obvious in that different holes are trapped a different number of times. Thus, if t_0 be the transit time of an untrapped hole, and t_t the average time a hole spends in a trap, the transit time of a hole trapped n times will be given by the equation

$$t_n = t_0 + nt_t. \qquad \ldots\ldots(2)$$

Assuming, then, that the holes which go to make up the peak of hole arrival pulse have all been trapped an equal number of times, we can arrive at a trap depth from the following argument. We can express the time in traps as

$$t_t = (1/B_0)\exp(E/kT) \qquad \ldots\ldots(3)$$

where B_0 is the relaxation frequency of a trap, expected to be about 10^9 sec^{-1} in germanium (see Torrey and Whitmer 1948), and E is the trap depth. Substituting for t_t in (2) and dividing by t_0 we have

$$t_n/t_0 = 1 + (n/B_0 t_0)\exp(E/kT). \qquad \ldots\ldots(4)$$

Then, if μ_{eff} is the effective mobility, μ_0 the mobility in the absence of trapping, L the distance from emitter to collector and F the sweeping field, then $t_n = L/F\mu_{eff}$ and $t_0 = L/F\mu_0$. Substituting for t_n and t_0 in (4), $\mu_0/\mu_{eff} = 1 + (n\mu_0 F/B_0 L)\exp(E/kT)$. At sufficiently low temperatures the exponential term becomes greater than unity, and

$$\mu_{eff} = (B_0 L/nF)\exp(-E/kt). \qquad \ldots\ldots(5)$$

Thus the trap depth may be calculated by plotting $\ln \mu_{eff}$ as a function of $1/T$. By this method a trap depth of 0·11 ev was determined.

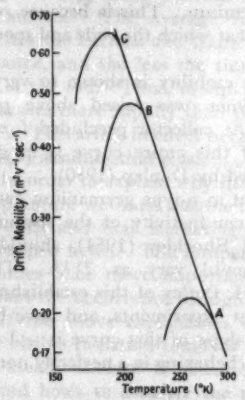


Figure 4. Drift mobility of injected holes as a function of temperature showing typical trapping curves. A, aged $i_e = 0{\cdot}3$ mA; B, freshly etched $i_e = 0{\cdot}3$ mA; C, freshly etched $i_e = 0{\cdot}6$ mA.

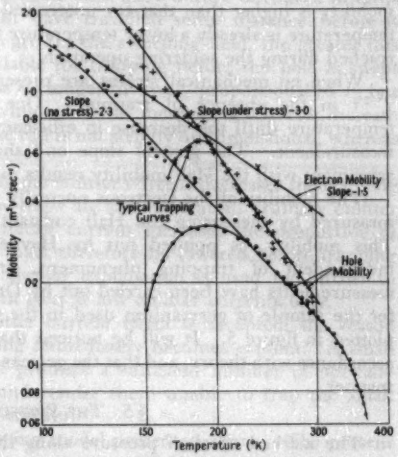


Figure 5. Collected results of mobility as a function of temperature.

The temperature at which trapping sets in is dependent upon the following parameters:

(*a*) Emitter current. The trapping temperature can be decreased by increasing the emitter current. Curve B of figure 4 was measured using 0·3 mA emitter current, while curve C was obtained simultaneously by using 0·6 mA.

(*b*) Sweeping field. It was found that the trapping temperature decreased with increasing sweeping field in much the same way as the emitter current effect.

(*c*) Surface condition. The trapping temperature for a given emitter current and sweeping field was never found to be constant for the same specimen from day to day. Whereas curve B of figure 4 was measured on a freshly etched rod, curve A was measured on the same rod under the same conditions of emitter current and sweeping field, but after the rod has been in contact with the atmosphere for some days.

(*d*) Illumination. The effects of trapping can be eliminated by illuminating the germanium with light from a tungsten bulb. This effect is important in that it has enabled the temperature variation of mobility due to lattice scattering alone to be measured down to temperatures much lower than those at which trapping normally predominates.

Figure 5 shows the collected results of a number of runs with different rods cut from the same sample of germanium. Some of the specimens had been subject to considerable stress due to differential expansion between the germanium and the Perspex on which they were rigidly mounted. For these specimens the drift mobility varied as T^{-3}. Even at room temperatures the mobility is higher than that for unstressed germanium. This is because room temperature is already a lower temperature than that which the rods and mounts reached during the soldering operation.

When no mechanical strains are present the mobility is shown to vary as $T^{-2\cdot3}$ in the absence of trapping. One specimen was raised above room temperature until the decrease in efficiency of the collector precluded further measurement. The general slope and shape of this entire curve is in good agreement with the Hall mobility results published by Dunlap (1950).

The mobility of the electrons normally present in n-type germanium can be measured by measuring the Hall constant and conductivity of the specimen. This mobility, as pointed out by Haynes and Shockley (1951), should be independent of trapping phenomena, and should vary as $T^{-1\cdot5}$. These measurements have been carried out by Dr. E. H. Putley of this establishment for the sample of germanium used in the present experiments, and have been plotted in figure 5. It will be noticed that the slope of this curve is $-1\cdot5$ in accordance with theory, and that the germanium is behaving in a perfectly normal manner.

§ 5. The Pressure Effect

The marked effect of pressure along the length of the germanium rod has been investigated at room temperature by loading a germanium rod with known weights. It was found that the drift mobility of injected holes increased linearly with applied thrust. For rods of the dimensions 1 cm × 0·05 cm × 0·05 cm the change in mobility is of the order 4% per kilogramme thrust. There is some evidence that the mobility decreases with extension, and that the resistivity also varies in much the same way as the mobility for both compression and extension.

§ 6. Discussion of Results

Although it has been established beyond reasonable doubt that some form of trapping agency can be responsible for the reduction of the mobility of injected

holes in germanium, it remains to see whether the observed effects in the trapping region can be explained on the basis of a conventional trapping model.

In the first place the variation of trapping temperature with surface treatment is strongly in favour of a model involving surface traps. If this be the case, the application of a magnetic field at right angles to the flow of holes should enhance trapping by deflecting a greater proportion of the holes towards the surface. A specimen was lowered in temperature until the mobility began to decrease due to trapping. The application of a transverse magnetic field reduced the mobility still further, no matter in which transverse direction the field was applied.

The placing of the traps on the surface now enables us to explain the sweeping field effect, namely that the trapping temperature decreases with increasing sweeping field. Reference to eqn (5) shows that although the field F is built into the equation, variation of F will not affect the effective mobility since the number of times n that a hole is trapped in transit down the rod is inversely proportional to F. This is because n is directly proportional to the time t_0 that a hole spends in the rod untrapped, and t_0 is inversely proportional to F.

We therefore build the sweeping field effect into the model in the following way. Holes leaving the emitter travel down the rod under the influence of the field F acting along the rod. At the same time they diffuse outwards towards the surface. The average hole will have travelled some distance before it reaches the surface and traps. The greater the sweeping field, the greater this distance, and the less the time available for trapping before the hole is swept out of the rod. Thus, in eqn (5) n is no longer inversely proportional to F, but decreases more rapidly than this, giving rise to an increase in μ_{eff} with increase in F. This arises in effect because the rod is not strictly unidimensional, whereas eqn (5) assumes that it is.

In order to explain why increasing the emitter current decreases the trapping temperature we must assume that the traps are saturated at all ordinary emitter currents, and that increasing the emitter current increases the proportion of untrapped holes. The temperature must therefore be lowered before trapping becomes once more important. A similar explanation is used by Sittner to explain the occurrence of high current gain at low emitter currents only.

If the above explanation of emitter current effect is accepted, the reason that visible radiation removes trapping immediately becomes evident. A small and steady amount of radiation will produce a sufficient number of optically excited holes to saturate the traps and render them unable to trap the small number of additional holes from the emitter.

There is an objection to the trapping model presented above. The present results have consistently yielded a trap depth of 0·11 ev, whereas Sittner's experiments on point type transistors give a room temperature trap depth of 0 31 ev. There are three possible explanations for this difference. Sittner found that his trap depth decreased with decreasing temperature, and explained this on the basis of traps of distributed depth. Since the value of 0·11 ev was obtained from the present experiments at a temperature of about 200°K, the explanation of distributed trap depth could well apply.

Secondly, Sittner applied filamentary transistor theory to the point contact transistor, a procedure which may not be justified.

The third explanation is that the trap depth actually varies with temperature. A new model consistent with this view will now be introduced. It has the

attraction that it introduces no further energy states than those already used in the conventional energy level diagram.

In figure 6 is shown the energy level diagram for 2 Ωcm n-type germanium. The surface barrier height ϕ will be of the order 0·4 ev at a temperature of about 200°K, and the barrier width W about 10^{-4} cm. The surface states S (see Bardeen 1947) have been placed at the top of the full band at the surface in accordance with the results of Hogarth (1953). A group of holes has been injected into the body of the germanium from the emitter. These holes will diffuse away from the emitter, eventually finding their way into the depletion layer W. When once in this region, they will 'gravitate' upwards towards the surface states, losing ϕ ev of energy to the lattice. In this condition many of the holes will be trapped in a part of the barrier narrower than the mean free path (about 10^{-6} cm) of a hole, and in consequence their mobility will be reduced

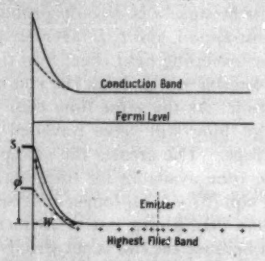


Figure 6. Energy level diagram for n-type germanium showing the surface potential barrier acting as a trap for holes.

considerably, in much the same way as has been suggested by Krömer (1953). The top of the barrier therefore acts as a low mobility 'channel' running the length of the germanium rod, and when a hole is found in this channel its mobility is so low that for all practical purposes it may be considered to be stationary in a trap equal in depth to the barrier height. As the number of holes in the channel is increased, the barrier height ϕ will be reduced, as shown by the dotted line, and fewer holes will be maintained in thermal equilibrium in the barrier.

We are now in a position to attempt an explanation of the observed effects on the basis of the new model. The emitter current effect now becomes fairly obvious. Any increase in the emitter current will lower the barrier as explained above, and fewer holes will be found in the very low mobility region: thus the average mobility will be increased.

We must here assume that the reason why the measured trap depth was 0·11 ev, and not 0·4 ev, the undisturbed barrier height, was that in order to obtain a reasonable signal at the collector, a large emitter current had to be applied, thus reducing the barrier height to a low value.

The effect of illumination, namely that trapping is reduced or eliminated, can be explained on the same basis as the emitter current effect. Hole–electron pairs are formed optically, many of them already in the barrier. These holes have the same effect upon the barrier as those injected by the emitter, and the

barrier may be reduced sufficiently to remove almost completely the low mobility region (cf. Hogarth 1953).

Variation of the trapping temperature with surface condition or age becomes a direct consequence of the variation of barrier height with surface condition, or surface state density.

Finally, we may utilize the explanation already given for the variation of the trapping temperature with sweeping field, except that we say that the effective length of the channel available for trapping varies rather than the number of traps.

There are some rather satisfying aspects of the trapping model as presented above. In the first place it explains the discrepancy between the trap depth as measured in this way, and secondly, it explains most of the observed trapping phenomena on the basis of the same mechanism, the variation of barrier height. Finally, it accounts for the inability of long wavelength light to empty the 'traps', an observation made by Navon, Bray and Fan (private communication from Dr. Bray) while investigating the lifetime of holes injected into germanium at low temperatures. They found that in addition to holes with a lifetime of some microseconds there also existed some with a lifetime very much greater than this. It was assumed that these holes had been trapped in traps which did not act as recombination centres. The reason that such traps cannot be emptied optically is that the optical transition ϕ ev within the full band is forbidden.

Acknowledgments

The author wishes to acknowledge the helpful guidance and encouragement of Dr. A. F. Gibson throughout this work. He is also indebted to his colleagues at Telecommunications Research Establishment for much helpful discussion.

This paper is published by permission of the Chief Scientist, Ministry of Supply, the Controller, H.M. Stationery Office, and the Senior Representative, Australian Department of Supply.

References

Bardeen, J., 1947, *Phys. Rev.*, **71**, 717.
Dunlap, W. C., 1950, *Phys. Rev.*, **79**, 286.
Haynes, J. R., and Shockley, W., 1951, *Phys. Rev.*, **81**, 835.
Hogarth, C. A., 1953, *Proc. Phys. Soc.*, **66**, 845.
Hunter, L. P., Fleisher, H., and Irish, C. G., 1952, *Trans. Inst. Radio Engrs., N.Y.* Professional Group on Electron Devices No. 1, p. 25.
Krömer, H., 1953, *Z. Phys.*, **134**, 435.
Lawrance, R., 1953, *Phys. Rev.*, **89**, 1295.
Lawrance, R., and Gibson, A. F., 1952, *Proc. Phys. Soc.* B, **65**, 994.
Mitchell, W. H., 1954, *J. Sci. Instrum.*, in the press.
Shockley, W., 1950 a, *Electrons and Holes in Semiconductors* (New York : Van Nostrand); 1950 b, *Phys. Rev.*, **78**, 294.
Sittner, W. R., 1952, *Proc. Inst. Radio Engrs., N.Y.*, **40**, 448.
Torrey, H. C., and Whitmer, C. A., 1948, *Crystal Rectifiers* (New York : McGraw-Hill).
Valdes, L. B., 1952, *Proc. Inst. Radio Engrs.*, N.Y., **40**, 445.

The Magnetic Susceptibility of Metallic Uranium

By L. F. BATES and D. HUGHES
Department of Physics, University of Nottingham

MS. received 16*th September* 1953

Abstract. The magnetic susceptibility of a pure specimen of metallic uranium has been measured over the temperature range 15° to 920°C. It is proved that sudden changes in susceptibility accompany the α–β and β–γ phase changes at 660°C and 767°C respectively. The results are compared with those of other workers, and the difficulties of providing an adequate explanation of the thermomagnetic behaviour are outlined.

§ 1. Introduction

Since the second world war three sets of experiments on the magnetic properties of uranium have been reported. It is known that the metal phase stable at room temperature, the α-phase, is orthorhombic and, according to Jacob and Warren (1937), may be considered to be a distorted hexagonal close-packed lattice in which four of the twelve nearest neighbours have moved still closer to one another than would be the case in an undistorted lattice. According to Wilson and Rundle (1949) the α-phase persists up to 640°C only, but Tucker (1951) gives the temperature at which it transforms to the β-phase as 660°C. The β-phase in turn transforms to the γ-phase at 760°C. Tucker states that the β-structure is tetragonal with 30 atoms per unit cell, while the γ-structure is body-centred cubic. It is clearly desirable to have accurate magnetic measurements to determine whether the above phase changes are accompanied by corresponding changes in susceptibility at clearly defined temperatures.

An examination of the magnetic behaviour of a fairly pure specimen of uranium, stated to contain some 700 parts in 10^6 of iron, 60 in 10^6 of nickel and 5 in 10^6 of cobalt, was made by Bates and Mallard (1950). They used a modified Curie method, with special pole-pieces and electromagnetic dynamometer control, and found the susceptibility at 20°C to be $+1{\cdot}740 \times 10^{-6}$ e.m.u. g^{-1}, increasing to $+1{\cdot}804 \times 10^{-6}$ at 350°C. They also found that alloys of pure uranium with 5 and 10 atomic % of iron showed as did the pure uranium no traces of ferromagnetism, at any rate in the moderate fields at their disposal. Assuming that the iron was all present in the form of the compound U_6Fe, they computed that the latter substance had a susceptibility of $+2{\cdot}036 \times 10^{-6}$ e.m.u. g^{-1} at 20°C and that its susceptibility remained practically independent of temperature over the stated range.

Bates and Mallard (cf. Bates 1952) later reported measurements on these and other samples of pure uranium over the temperature range -183°C to 18°C. They found that all the specimens were characterized by a pronounced rise in susceptibility at lower temperatures, some specimens exhibiting a definite minimum in the susceptibility–temperature curve. They found that a specimen of U_6Fe showed a very large increase in paramagnetic susceptibility below

-100°c, although with their experimental arrangements it could not be determined that this was due to an onset of ferromagnetism as distinct from a large increase in paramagnetism at the lower temperatures.

Turning now to measurements made at higher temperatures we have those by Kriessman and McGuire (1952), who used specimens of uranium, supplied by Messrs. Johnson, Matthey, which were of about the same iron content and impurity as the initial specimen used by Bates and Mallard. The first set of their measurements, termed A, showed the same relative rate of increase in susceptibility as found by Bates and Mallard, but the absolute value at room temperature was somewhat lower, viz. $1{\cdot}72 \times 10^{-6}$ e.m.u. g^{-1}, and they found a 4 parts in 10^6 ferromagnetic impurity at 25°c. They recorded an abrupt change in susceptibility at about 698°c which they associated with the α–β transformation. At 808°c a second and smaller abrupt change was recorded, and was attributed to the β–γ change. On cooling, a temperature hysteresis of the magnetic properties was observed. Kriessman and McGuire made a second set of measurements, B, after washing their specimen in nitric acid and resealing it in a fresh quartz capsule. The B set is in better accord over the relevant temperature range with the results of Bates and Mallard than is the A set. But the enclosure of uranium in quartz is in our view objectionable, because the metal may react with the quartz at high temperatures, and because of the obvious uncertainty of the temperature of the specimen as recorded by a thermojunction outside the quartz specimen tube.

Measurements were also made on four specimens, likewise sealed in quartz capsules, by Gordon and Kaufmann (1952) who recorded a room temperature susceptibility of $1{\cdot}66 \times 10^{-6}$, as compared with a mean value of $1{\cdot}74 \times 10^{-6}$ obtained by Bates and Mallard, and abrupt changes at 676 ± 10°c and 788 ± 10°c. Gordon and Kaufmann used a Bitter magnet with an iron core, measuring the magnetic forces with a chemical balance mounted in a chamber evacuated down to 10^{-4} mm Hg. Incidentally, in the course of their work UFe_2 was found to be ferromagnetic, whereas UMn_2 appeared to be antiferromagnetic, so that it may be possible to make ferromagnetic or antiferromagnetic ternary alloys according to the ratio of Fe to Mn. They found all metallic alloys of the type U_6Mn, U_6Fe, U_6Co and U_6Ni to have very nearly identical susceptibilities of approximately 2×10^{-6} e.m.u. g^{-1} at room temperature, increasing with rise in temperature as in the case of α-uranium but to less marked extents.

We may summarize the above results by stating that thus far uranium has been shown to be paramagnetic with an unusually large rate of increase of susceptibility with temperature; abrupt changes in susceptibility, presumably connected with α to β and β to γ transformations, have been recorded at markedly different temperatures by different sets of observers; temperature hysteresis is observed in the region of these temperatures; large increases in paramagnetism are found at liquid air temperatures; and, finally, uranium forms non-ferromagnetic alloys with iron and other ferromagnetic elements.

It is obviously a matter of some importance to know precisely the temperatures at which the magnetic susceptibility changes abruptly as uranium is heated. For this purpose we considered it necessary to have a furnace, capable of running continuously for many days, to permit a slow rate of change of temperature of approximately 1°c per minute to be arranged throughout a set of readings, and for the temperature of the specimen to be measured by means of a thermojunction

in direct contact with it. Because uranium oxidizes rapidly at high temperatures, and because of its known toxic properties, it is essential to heat the specimen *in vacuo* and to take rigorous safety precautions to avoid serious consequences from an accidental breakdown with ingress of air at high temperatures. With these points in mind the following arrangements were made.

§ 2. Experimental Arrangements

The magnetic measurements were made by a modified form of Gouy method using a Sucksmith ring balance with an electromagnet working under controlled conditions giving a field of about 9000 oersteds. Each uranium specimen was approximately 4 cm long, 4 mm in diameter and weighed about 9 g. A longer specimen would have been desirable, but the additional weight would have been most objectionable. The specimen S was mounted in a simple cradle of No. 42 s.w.g. tungsten wire and suspended from the ring balance as shown in figure 1(*a*), which is not drawn accurately to scale. It is desirable that the

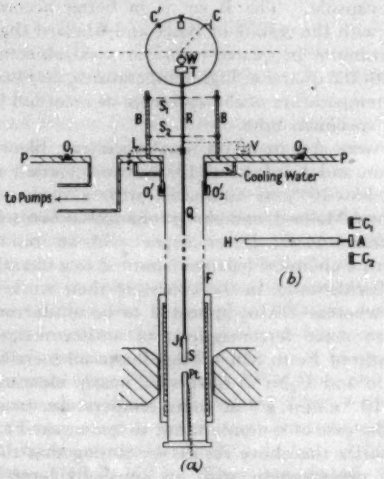


Figure 1. (*a*) diagram of apparatus, (*b*) calibration device.

suspension be as rigid as possible, and therefore a fine straight quartz tube Q, 0·5 mm in internal diameter and 1·0 mm in external diameter, was used. At its lower end was a quartz hook on which the tungsten cradle was hitched; this was the only non-rigid joint in the system, and apparently gave rise to no adverse effects. At the upper end of the tube was a small quartz collar and hook with which the tube Q was attached and cemented by Durofix glue to a small brass collar on the threaded brass rod R bolted to the downmost portion of the ring.

The temperature of the specimen was measured by a standard Pt–Pt 13% Rh thermocouple, whose hot junction J was in direct contact with the specimen. The thermocouple wires were suitably insulated electrically by taking the stiffer

Pt.Rh wire through the quartz tubing, while the more flexible Pt wire ran outside the latter, and was attached to it at intervals with touches of alundum cement. The temperature was determined from the calibration chart supplied by the makers (Baker Platinum, Ltd.). We observed no ill effects arising from the contact of the thermojunction with uranium, and we were advised that such effects were unlikely below 1000°c.

The ring balance was made from a strip of beryllium copper, Mallory 73, of cross section 0·1 mm × 3·0 mm, recommended by Professor Sucksmith, bolted to form a ring of 4·5 cm diameter. One concave mirror C of focal length 1 m was used instead of the customary two plane mirrors for convenience in setting up the apparatus, symmetry being conserved by a nut and bolt C′ set in the position normally occupied by a second mirror. For a scale distance of 1 m the balance had a sensitivity of 20 mm deflection per gramme load.

In an extended series of magnetic measurements the balance must be frequently calibrated. Consequently a V-shaped trough T was attached to the threaded brass rod R so that a ring-shaped mass W could be placed in T by remote control. W was moved by a pivoted brass rod with a hook H at one end to engage W and a piece of permanently magnetized alloy A at the other, as shown in figure 1 (*b*). Two small coils C_1 and C_2 wound on soft iron cores and connected in series were used to produce a magnetic field to move the alloy. There were two stable positions of the latter, namely, when it was held against either of the unexcited iron cores by its own magnetism. A directed pulse of current through the coils served to move the alloy from one core to the other, and so place W in, or remove it from, the trough T. This device worked well and permitted calibrations without impairing the vacuum or having a permanent current producing heating effects near the ring. The scale deflections were proved to be directly proportional to load by measuring deflections for a given field, both when the weight was on the balance and when it was not. This test would perhaps not normally be necessary, but in these experiments the balance was rather more severely loaded and distorted than is normally the case.

A beam of light from a 60 watt lamp with a clear bulb passed through a rectangular housing aperture of 2·4 cm × 1·3 cm to be reflected from the mirror C. The reflected beam fell upon a split selenium photoelectric cell whose halves were connected in opposition through a galvanometer, zero reading indicating the position of the spot. The cell was mounted on a micrometer which could be turned by a remote control system of Magslip servo-motors (Muirhead and Co. Ltd., Bulletin B-580-C). Since the photoelectric cell was used in purely null measurements, its linearity of response, fatigue, etc., did not come into question, although it was nevertheless frequently tested. The sensitivity was such that around the null setting a movement of the micrometer head by 2 divisions (0·02 mm) corresponded to 1 division (approximately 2 mm) on the (spot) galvanometer scale. The stability of the system was closer than $\pm 0{\cdot}5\%$ and the micrometer could be set and read within this limit.

To prevent sideways motion of the specimen there were two flat spirals S_1, S_2 bolted to two threaded brass supports BB. Each spiral consisted of a two and a half turn spiral slit $\frac{1}{16}$ in. wide cut in a flat sheet of beryllium copper 8 cm square and 0·1 mm in thickness. The spirals were fixed some 4·5 cm apart. The system was oil damped by a flat vane V attached to the rod R as shown, dipping in apiezon B vacuum oil. The elastic properties of the ring were maintained

constant by the shielding from direct radiation from the furnace provided by the spirals S_1, S_2 and by water cooling the base PP. The balance was covered with a glass dome, provided with side tube with window, which rested on an O-ring seal inset in the plate PP, and shown at O_1 and O_2.

The furnace consisted of a transparent quartz tube F some 61 cm long and 2·0 cm external diameter with wall thickness 2 mm. This tube was wound over a length of 26 cm with No. 32 s.w.g. platinum 20% rhodium wire. The upper end of F was connected to the plate PP with a readily demountable O-ring seal, O_1' O_2', as shown in figure 1 (*a*), while its lower end was fitted with a Tufnol plug. To reduce temperature gradients in the neighbourhood of the specimen, a cylindrical molybdenum sheath some 0·3 mm thick and 15 cm long lined the central portion of F. The furnace winding was coated with alundum cement and slipped inside a double walled brass tube whose inner wall was of circular and outer wall of rectangular section which formed a water jacket. The whole was clamped between the water-cooled pole pieces of an electromagnet, the position of the specimen with respect to the (circular) pole pieces being approximately as shown in figure 1 (*a*).

As the furnace often ran continuously for many days it was necessary to stabilize the temperature within it. A platinum resistance thermometer Pt was made by winding a 10 ohm coil non-inductively on the end of an aluminous porcelain tube, supported from the Tufnol plug through which copper leads to the platinum were brought through waxed holes. The resistance thermometer formed one arm of a Wheatstone bridge circuit with ratio arms 150 ohm, energized by a valve oscillator at about 310 c/s. The remaining arm consisted of a radio-type non-inductive rheostat and some non-inductive fixed resistors to permit a continuous range of temperature control up to 1000°c. The out-of-balance voltage on the bridge was applied to three stages of valve amplification before being delivered to the control grid of a pentode in a gating circuit. The amplified signal was here compared in phase with a signal of constant phase applied to the suppressor grid of the pentode from the valve oscillator activating the bridge. In this way the gating system discriminated between out-of-balance voltages due respectively to the resistance thermometer being at a temperature higher or lower than that appropriate to the setting of the control resistance. The output from the gating stage—large when the compared signals were in phase and small when they were out of phase—was fed to a power amplifier and Sunvic vacuum relay. The latter switched the furnace current between a high value and a low one; actually, the latter was zero for furnace temperatures below 300°c. These currents were fixed by chosen settings on rheostats connected in the relay circuit, and the temperature inside F was so maintained constant to ± 1°c during a determination of the magnetic susceptibility.

The pumping system was a two-stage backing pump and two-stage mercury-in-glass diffusion pump, with a liquid air vapour trap which required filling once in 24 hours. A 40 mm diameter glass pumping line was joined to the underside of the plate PP with a flexible metal bellows; an ionization pressure gauge was incorporated, and registered an ultimate pressure of 3 to 4×10^{-5} mm Hg.

Because it was felt that slow structure changes were probable with uranium, it was decided to make experiments over a long period of time, and to raise the temperature gradually from one chosen point to another. Actually the run in

which our final measurements were made lasted 10 days. It was therefore essential to use safety devices to avoid troubles due to failure of power, vacuum or water supply. The cooling water for the furnace passed first through the diffusion pump condenser, and eventually passed to the waste outflow through a glass jet. A pressure difference was therefore established across the jet and this was registered on a mercury manometer. The column of mercury bridged two electrical contacts when the water supply was satisfactory and so operated an electromagnetic relay. If, however, the water supply fell the mercury column also fell and the relay was opened. The latter action caused the heater of the diffusion pump to be switched off and also caused a Sunvic vacuum relay to open and so break the furnace current.

If the vacuum failed, the ionization gauge and furnace currents were switched off in the following manner. The pressure was normally registered as a reading on a 0–1 milliammeter in the ionization gauge equipment. A meter of similar type was connected in series with that already in the equipment, and arranged so that its needle forced a fine platinum wire into a cup of mercury when the ionization gauge current rose above 0·8 ma. When this happened, a large negative bias was applied to the grid of a power pentode with two Sunvic relays in its anode circuit. The relays were thereupon opened and the furnace and ionization gauge currents were broken. The furnace and electromagnet currents were supplied from the d.c. mains, and when these failed no harm could result. The backing pump and the diffusion pump were worked from the a.c. mains supply and, if the latter failed, the two Sunvic relays mentioned above cut out, and the pumps and furnace were switched off. On restoration of the mains supply only the backing pump automatically restarted, while the other circuits remained dead.

One final point should be mentioned. We only obtained reliable results with our apparatus when it was mounted on a concrete floor in the basement of the building. On the second floor of a well-built laboratory there was sufficient floor movement to upset the zero of the balance continuously.

§ 3. Experimental Results

The uranium specimens were kindly supplied by the Atomic Energy Research Establishment, Harwell. The present experiments were confined to two specimens. The first originally showed evidence of preferred crystal orientation, and in the course of preliminary work, during which the apparatus was continually being improved, this specimen recrystallized in a remarkable way, giving long crystals running parallel to the axis of the specimen, and the whole specimen became considerably curved because of unequal crystal growth. A similar phenomenon, but to a far less extent, was observed with the second specimen. However, the results obtained with the first specimen were of value in that they were entirely consistent with the results obtained with the second specimen.

Chemical and spectroscopic analysis of the latter showed that it was very pure and probably contained less than 50 and 20 parts per million of iron and silicon respectively. From our point of view quantities of chromium and molybdenum would be important because they are said by Wilson and Rundle (1949) to depress the temperatures at which the several phase transitions occur. It was important to know whether the ferromagnetic impurities were present

in the free state or in the form of ferromagnetic compounds, and separate measurements were made to determine the magnitude of any ferromagnetic effects by the Honda and Owen method (cf. Bates 1951). This test requires much larger fields than were employed in the main susceptibility measurements, because if free iron, for example, were present in the form of tiny spherical particles, the demagnetizing field acting on each such particle would be $4\pi I/3$ and, consequently, the particle would behave only as a strong paramagnetic until its maximum demagnetizing field $4\pi I_{max}/3$ was exceeded.

The second specimen was therefore suspended for these tests from the ring balance with the furnace removed and the pole pieces of the magnet brought much nearer together. The fields acting on the top and bottom of the specimen were measured by a small search coil and fluxmeter, using a mechanical device for extracting the search coil quickly from the magnet field. Fortunately, the value of the field at the upper end of the specimen was small, and its presence involved a correction of only 1%. The susceptibility was measured at a series of eight values of the lower field H between 12 500 and 17 350 oersteds. A horizontal straight line was obtained on plotting the susceptibility against $1/H$. Consequently, ferromagnetic effects were absent.

The complete set of 76 measurements on the second specimen are shown in figure 2, where the circles refer to readings taken while the specimen was being raised from room temperature to the highest temperature, 920°c, and the crosses refer to readings taken during the descent to room temperature; each point is

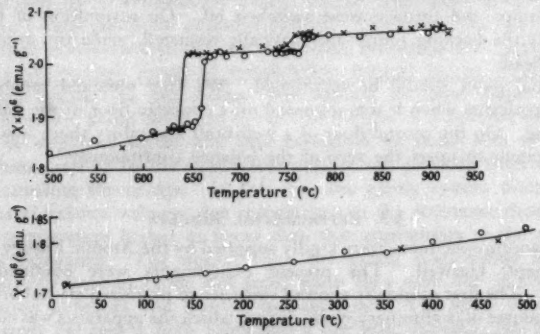


Figure 2. Change of susceptibility of uranium with temperature.

a mean of at least two readings differing by less than 0·5%. A few representative results are given in table 1 to illustrate more clearly the numerical behaviour. Incidentally, by the Gouy method one actually measures the volume susceptibility multiplied by the area of cross section of the specimen. As the cross section increases as $1+2\alpha T$ for an increase of temperature T°c and the density decreases as $(1-3\alpha T)$, our results should each be multiplied by a factor $(1+\alpha T)$, where α is the mean coefficient of linear expansion of uranium. However, the maximum correction, applicable to the result for the highest temperature, could not exceed 2%; this correction has therefore not been made.

Since we used a modified Gouy method the actual susceptibility measurements were relative to the room temperature value, which had to be found separately. Consequently the absolute value at 20°c was kindly found for us by Mr. S. J. Leach, using an apparatus similar to that of Bates and Mallard, after completion of the main work. This was not as accurate as could be desired because the specimen had to be cut and mounted in a Perspex guide, and its slight curvature, which was unimportant in the main work, now caused difficulty. The value obtained by Mr. Leach for χ_{20} was $1{\cdot}70 \pm 0{\cdot}03 \times 10^{-6}$ e.m.u. g^{-1}, and we have taken $1{\cdot}72 \times 10^{-6}$ e.m.u. g^{-1} as the basic value for the results given in table 1 and the graph.

Table 1

Temp (°c)	$\chi \times 10^6$	Temp (°c)	$\chi \times 10^6$	Temp (°c)	$\chi \times 10^6$
14	1·720	770	2·049	648	2·021
205	1·757	916	2·069	640	1·876
353	1·786	921	2·058	368	1·791
654	1·891	761	2·056	124	1·743
671	2·015	751	2·037	15	1·720
764	2·021				

χ stated in e.m.u. g^{-1}.
(See text regarding accuracy of absolute and relative values.)

§ 4. Discussion of Results

We think that discussion of the results is best facilitated by referring to table 2. In the first main column of this table are given references to observers or the type of measurement on which information given in the table is based. In the second main column are given the values of χ_{20} obtained by different workers; we have already commented on the limits of accuracy of our own value.

Table 2

Source	χ at 20°c e.m.u. $g^{-1} \times 10^6$	Mean $\frac{1}{\chi}\frac{d\chi}{dT} \times 10^4$			Transition temp (°c)				$(\delta\chi/\chi_T) \times 100$	
		α	β	γ	α–β	β–γ	γ–β	β–α	α–β	β–γ
(1)	1·72 ± 0·03	1·4	zero	(0·35)	660	767	758	647	7·2	1·4
(2)	1·74 ± 0·003	(1·2)	—	—	—	—	—	—	—	—
(3)	1·66 ± 0·06	1·2	zero	0·95	676	788	—	—	8·5	2·5
(4)	1·72	1·8	zero	0·59	698	808	783	658	6	1·5
(5)	—	—	—	—	660	760	—	—	—	—
(6)	—	—	—	—	640	760	—	—	—	—
(7)	—	—	—	—	667	772	764	645	—	—
(8)	—	—	—	—	661	770	—	—	—	—

Notes: (1) Bates and Hughes (Leach),
(2) Bates and Mallard (1950),
(3) Gordon (and Kaufmann) (Thesis) (1952),
(4) Kriessman and McGuire (1952),
(5) Crystal structure, Tucker (1951),
(6) Crystal structure, Wilson and Rundle (1949),
(7) Electrical resistance, Dahl and van Dusen (1947),
(8) Specific heat, see table 1, Kriessman and McGuire (1952).

In the third main column are given the mean values of $\chi^{-1}d\chi/dT$ or the temperature coefficient of increase of susceptibility over the maximum temperature range possible for the several phases in the experiments. The value for the α-phase given against the names of Bates and Mallard is placed in brackets to indicate that it was based on measurements ranging from 20° to 350°C only, whereas the other values in this sub-column were based on measurements from 20° to the high temperature at which the α-phase was assumed to disappear. Our value for the γ-phase is also in brackets to indicate that the temperature range was not as extensive as that used by the other workers.

In the fourth main column are given the several transition temperatures as recorded by the different observers. A good idea of the magnitude of the temperature hysteresis may be obtained from the table and from the graph of figure 2. Each recorded magnitude presumably depends to some extent on the rate of heating and cooling during the experiments.

The final main column gives the magnitudes of the abrupt changes of susceptibility at the transition points; they are given as $100(\delta\chi/\chi_T)$ where $\delta\chi$ is the observed sudden change in susceptibility and χ_T is the absolute value of the susceptibility at the beginning of the phase change. On the basis of our experiments we can definitely state that the abrupt magnetic changes occur at precisely those temperatures suggested by metallurgical evidence. The differences between the latter temperatures and those recorded in the magnetic experiments of other workers are undoubtedly due to the fact that the uranium was enclosed in quartz and the thermojunction was not directly in contact with the metal during their magnetic measurements.

We have at present no completely satisfactory explanation to offer for the marked rise in susceptibility with temperature between −100 and 660°C We do not get much help from Stoner's free electron theory (Stoner 1935, 1936) for this requires a paramagnetism which decreases with temperature, although by taking account of the coefficient of cubical expansion of uranium, which has the values 48×10^{-6} and 69×10^{-6} per deg C over the ranges 20 to 100°C and 300 to 400°C respectively, the Stoner theory might be made to predict a small increase of susceptibility with temperature, but it would certainly be much smaller than the experimental value.

The trouble is that the term in T^2, which gives the variation of susceptibility with temperature in Stoner's theory, is negative for an isolated electron energy band. However, it has been suggested by Wohlfarth (1949), in explaining the increase of susceptibility with temperature of the copper-rich alloys of the copper–nickel series, that two energy bands may overlap and so produce a positive T^2 term with a corresponding increase of susceptibility with temperature. The effect of the overlap may be looked upon as arising from a transfer of electrons between the two bands. Increase of temperature causes an increase in the number of electrons at the bottom of the upper energy band with an increase of the number of positive holes in the lower energy band, and so might overcome the normal temperature decrease in susceptibility which would be found with two separated bands.

Although the appearance of figure 2 may suggest that up to 650°C uranium may be an antiferromagnetic substance, the presence of an α–β phase change would seem to rule it out. Moreover, Mallard (cf. Bates 1952) found that between room temperature and liquid air every specimen of uranium which he

examined showed a marked increase in susceptibility below $-100°$C, a region where another phase change in uranium has not been reported. The low-temperature behaviour seems to depend very much on the purity of the specimen. It may be noted that Hoare and Matthews (1952) have recently discovered an unexpected peak in the susceptibility temperature curve for palladium at about 80°K, and they find difficulty in accounting for this peak either in terms of a collective electron theory or in terms of antiferromagnetism, but Rhodes (cf. Hoare 1953) states that if band forms are properly chosen such peaks can be explained.

The more or less constant paramagnetic susceptibilities of the β- and γ-phases are interesting. One can readily postulate a slight increase in the number of free electrons to account for the sudden increases at the transition points, but one cannot understand why the susceptibilities do not decrease with rise in temperature unless thermal expansion provides adequate explanation. Further experiments are being made on uranium alloys at high temperatures in an attempt to provide additional information on these points.

Acknowledgments

Our warm thanks are due to the Director, Atomic Energy Research Establishment, for permission to publish the results. We should also like to express our gratitude to the members of the Metallurgy Division of A.E.R.E. for their advice and help on many problems.

References

Bates, L. F., 1951, *Modern Magnetism* (Cambridge: University Press); 1952, *The Transition Metals*, Ismet 1048/BP 169.
Bates, L. F., and Mallard, J. R., 1950, *Proc. Phys. Soc.* B, **63**, 520.
Dahl, A. I., and van Dusen, M. S., 1947, *J. Res. Nat. Bur. Stand.*, **39**, 53.
Gordon, P., 1952, *Thesis*, Massachusetts Institute of Technology, A.E.C.U., 1833.
Hoare, F. E., 1953, *Nature, Lond.*, **171**, 1058.
Hoare, F. E., and Matthews, J. C., 1952, *Proc. Roy. Soc.* A, **212**, 137.
Jacob, C. W., and Warren, B. E., 1937, *J. Amer. Chem. Soc.*, **59**, 2588.
Kriessman, C. J., Jr., and McGuire, T. R., 1952, *Phys. Rev.*, **85**, 71.
Stoner, E. C., 1935, *Proc. Roy. Soc.* A, **152**, 672; 1936, *Ibid.*, **154**, 656.
Tucker, C. W., Jr., 1951, *Acta Cryst., Camb.*, **4**, 425.
Wilson, A. S., and Rundle, R. E., 1949, *Acta Cryst., Camb.*, **2**, 126.
Wohlfarth, E. P., 1949, *Proc. Roy. Soc.* A, **195**, 454.

Intrinsic Magnetization in Platinum Cobalt Alloys

By A. W. SIMPSON and R. H. TREDGOLD
Department of Physics, University of Nottingham

Communicated by L. F. Bates; MS. received 2nd October 1953

Abstract. A study of the effect of heat treatment on the intrinsic magnetization exhibited by alloys of platinum and cobalt containing small percentages of the latter material is described. The results are compared with those obtained from the platinum nickel series of alloys. Tentative conclusions concerning the nature of the exchange interaction in the two series of alloys are given.

§1. Introduction

Studies of the properties of alloys of ferromagnetic transition metals with non-ferromagnetic transition metals have been carried out by a number of investigators to obtain information which might be helpful in interpreting the properties of ferromagnetic metals. Particular interest attaches to investigations of binary alloys of metals where the d shells contain similar numbers of electrons, as in these investigations one relevant parameter, the degree of filling of the d band, is maintained at least approximately constant.

A study of the system Pt–Ni was made by Kussmann and Nitka (1938) and a study of the system Pt–Co by Constant (1930). Subsequent to Constant's work it was discovered by Newkirk, Smoluchowski, Geisler and Martin (1951) that the system Pt–Co shows order–disorder reactions around the compositions Pt 50, Co 50, and Pt 75, Co 25 atomic per cent. We considered that a study of the effects of ordering on the magnetic properties of Co–Pt alloys containing only small percentages of cobalt might yield interesting results, and accordingly decided to investigate the variation of intrinsic magnetization as a function of temperature for a series of alloys containing from 13 to 25 atomic per cent cobalt, and to make measurements on quenched and annealed specimens in each case.

§2. The Preparation of the Specimens

Some of the specimens used were supplied by Messrs. Johnson, Matthey and Company and some were lent by the Mond Nickel Company, the compositions being accurately known in every case. Each was cut to a size of $1\frac{1}{2}$ cm $\times \frac{1}{2}$ cm $\times \frac{1}{2}$ cm and annealed prior to subsequent heat treatment. In order to achieve the disordered state each specimen was heated to 1200°c and water quenched. After measurements had been made on it in this condition it was heated to 800°c and then cooled at a uniform rate of 2·5° per hour to a temperature of 400°c and finally furnace cooled from this temperature. This annealing process was carried out *in vacuo* in a temperature-stabilized mechanically controlled furnace.

§3. Method of Measurement

The measurements were carried out by a method based on the well-known technique devised by Weiss and Forrer (1926, 1929; cf. Bates 1951). The specimen was enclosed in a small furnace of platinum wire wound on a meerschaum former, a thermocouple being in contact with the specimen. The assembly was placed between the pole-pieces of an electromagnet and could be withdrawn rapidly using a mechanical system with remote control. Small coils placed on the extremities of the pole-pieces of the magnet were connected to a ballistic galvanometer. The deflection of the galvanometer on the withdrawal of the specimen was thus proportional to the magnetic moment of the latter. In order to calibrate the system and to correct for the effects produced by magnetic images the specimen was replaced by a water-cooled coil of known moment.

When making measurements above room temperature the furnace was enclosed in a small water-cooled container which was integral with and moved with the furnace assembly. In making measurements below room temperature the furnace was surrounded by a thin brass tube which remained stationary

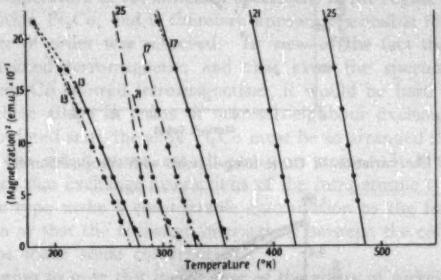


Figure 1. (Magnetization)^2 plotted against absolute temperature. The broken lines refer to annealed alloys. The atomic percentage of cobalt in each alloy is indicated beside the curve referring to it.

and which was immersed in a Dewar flask containing liquid air. It was thus possible to make measurements of intrinsic magnetization in the range -180°c to 400°c. The method of obtaining the variation of intrinsic magnetization with a temperature for $H=0$ was again based on the method employed by Weiss and Forrer (1926). At each chosen temperature the magnetization was measured for a series of magnetic fields ranging from 4000 to 9000 oersteds. Curves of I (magnetization) against T for H constant were then plotted. From these, curves of H against T for I constant were obtained and the value of T corresponding to $H=0$ was thus found for each chosen value of I. Finally, in order to ascertain the Curie temperature, I^2 was plotted against T for each specimen. As will be seen from figure 1, the latter curves became straight lines in the region immediately below the Curie temperature, and it was then possible, by taking the value of T corresponding to $I^2=0$, to determine the Curie point with precision. The values so obtained are estimated to be accurate to within ± 1°c.

As a check on the accuracy both of the composition of the specimens and of the experimental method it is of interest to note the Curie temperatures of two specimens, obtained from the two different sources quoted above and stated to be of the same composition and given identical heat treatment, agreed to within 1°c.

§4. Results

The variation of Curie temperature with compositon for both the quenched and annealed specimens is shown in figure 2. The values obtained by Constant (1930) on furnace-cooled specimens are shown for comparison. The rather surprising fact that the values obtained by him lie above the values obtained in the present work for quenched specimens may be explained as

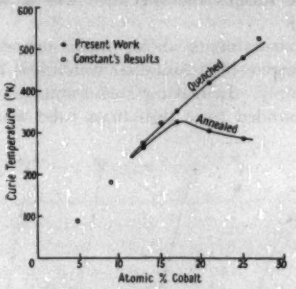


Figure 2. The variation of Curie temperature with composition for Pt–Co alloys.

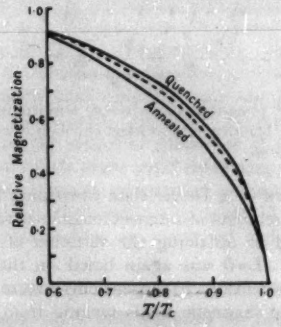


Figure 3. Intrinsic magnetization plotted against reduced temperature for the 25% Co alloy. The molecular field theory curve is shown dotted for comparison.

follows. Constant made measurements using fields of less than 1000 oersteds. It was found in the present work that for the majority of the specimens it was necessary to employ fields in excess of 4000 oersteds to obtain technical saturation. A certain lack of precision in Constant's results is thus to be expected. It should be pointed out here that, owing to the low values of magnetization of these alloys and hence low demagnetizing field, the distinction between internal field and applied field is trivial.

Figure 3 shows part of the (magnetization, reduced temperature) curves for the alloy containing 25 atomic per cent cobalt together with the $S=\frac{1}{2}$ curve of the Heisenberg first approximation and molecular field theory. The curve for the ordered alloy lies below the $S=\frac{1}{2}$ curve and the curve for the disordered alloy above it. This behaviour is the reverse of that associated with the order–disorder phenomena shown by the alloy Ni_3Fe and studied by Went (1951). Went suggested that the behaviour of Ni_3Fe could be explained as follows. The disordered specimen consists of a mixture of alloys of differing compositions and different Curie temperatures. The superposition of the (magnetization, temperature) curves for these alloys therefore produces a curve less concave towards the origin than the curve for the ordered, homogeneous alloy. This theory should apply to any ferromagnetic alloy possessing an order–disorder transition. The fact that Pt–Co does not show this behaviour indicates that, in this case at least, some other explanation is necessary.

§ 5. Discussion

Annealing the specimens at various different rates had only a small effect on the Curie temperature of the annealed specimens in the region corresponding to the composition Pt_3Co, and it therefore appears reasonable to assume that a fairly high state of order was achieved. In view of the fact that the ordered specimens remained ferromagnetic, and that even the specimen containing only 13 per cent Co showed ferromagnetism, it would be hard to explain the behaviour of these alloys in terms of nearest neighbour exchange interactions only. In the ordered state the alloy Pt_3Co must be so arranged that the nearest neighbours of the cobalt atoms are all platinum atoms. It is thus necessary to assume either that exchange interactions of the intra-atomic type as opposed to inter-atomic type make a considerable contribution to the forces favouring ferromagnetism or that the exchange interactions between the cobalt atoms and platinum atoms make some contribution.

It is interesting to note that in the case of the alloys of nickel and platinum studied by Kussman and Nitka (1938) an extrapolation from their results shows that alloys containing more than 65 atomic % platinum should be non-ferromagnetic. If it is assumed that in nickel the intra-atomic exchange integral is much less important than in the case of cobalt it is possible to explain the difference in behaviour of the two series of alloys. This explanation would also be consistent with the fact that nickel has a much lower Curie temperature than has cobalt.

Acknowledgments

The authors wish to express their thanks to Professor L. F. Bates for his interest and encouragement and for the facilities of his laboratories, and to Dr. L. B. Pfeil and the Mond Nickel Company for the kind loan of specimens.

References

Bates, L. F., 1951, *Modern Magnetism* (Cambridge : University Press), p. 287.
Constant, F. W., 1930, *Phys. Rev.*, **36**, 1654.
Kussmann, A., and Nitka, H., 1938, *Phys. Z.*, **39**, 373.
Newkirk, J. B., Smoluchowski, R., Geisler, A. H., and Martin, D. L., 1951, *J. Appl. Phys.*, **22**, 290.
Weiss, P., and Forrer, R., 1926, *Ann. Phys., Paris*, **5**, 153 ; 1929, *Ibid.*, **12**, 279.
Went, J. J., 1951, *Physica*, **17**, 596.

The Temperature Decay Law of a Naturally Convected Air Stream

By W. RAILSTON

Physics Branch, Royal Military College of Science, Shrivenham, Swindon, Wilts.

MS. received 21st August 1953

Abstract. Temperature measurements have been made both vertically and transversely in the plume of air rising by natural convection from an electrically heated gauze in a still atmosphere. Schlieren methods made the air stream visible, so that the type of air flow could be observed at different levels. The results are in good agreement with the dynamical theory of naturally convected gases developed by O. G. Sutton.

§1. Introduction

The mathematical analysis of convection problems being notoriously difficult, it was desired to make careful experimental measurements of the way in which the mean temperature of a hot air stream rising under convection falls off with increasing distance from the heat source, when in a still atmosphere. The maximum measured rise of temperature was kept small ($<21°$c) so that the density of the hot air and its surroundings could be considered sensibly constant.

The plan of the research divided itself into the investigation and use of the best methods of producing a uniform, stable hot air jet, and of measuring the temperature at a number of points across the jet, at each of many different distances above the heat source. Since theoretical analysis of this problem usually refers to a point source of heat, it was decided to use as small a source as practicable and then to make a theoretical extension to a true point source. In this way the variation of the mean temperature of the jet as a function of its height above a point source was found.

Little experimental work had been done on this problem, except by Schmidt (1941), who, unfortunately, gave few practical details of his work, especially of the source of heat, and omitted to state the rate of heat output from it. It will be shown later that, in making no attempt to render the air stream visible, he overlooked a valuable means of observing the type of air flow and of ensuring similar conditions throughout the experiment.

It was further decided to use several different heat outputs, so that a test could be made of the effect of the strength of the source on the temperature rise it caused at any given point.

§2. Experimental Procedure

Trials were made of different types of electrically heated resistance, aiming to produce a uniform, stable air jet. The best source used was a piece of fine mesh wire gauze, originally produced for the electrical heating of flying suits. A piece of copper wire was soldered to each of two opposite sides of a square of the gauze of 3·5 cm side. Alternating current flowed between these two sides, supplied from a variable auto-transformer receiving power from the

mains through a constant voltage transformer. The currents passed through the heater, together with the values of its resistance at the different temperatures they produced, were each measured on a precision potentiometer.

For temperature measurement at different points in the air jet, the platinum resistance thermometer, the thermocouple and the thermistor were each considered. The first was not used since, although special methods allow the use of only a few millimetres of platinum wire, it is still not suitable for measuring temperature at a point.

Copper–constantan thermocouples were used in some of the initial measurements, but some changes of calibration were noted. As the hot junctions were being frequently moved in the air jet, these may have been due to mechanical changes in the leads caused by bending or kinking. Because of this, and for greater convenience, thermistors were substituted for thermocouples. Thermistors themselves suffer from change of calibration, the resistance at a given temperature at first increasing slowly with time. This change was avoided by using thermistors which were at least a year old and which had reached a stable calibration by natural ageing. Artificial ageing, by heating to the maximum temperature to be measured for about 24 hours, was also successfully used.

The connection between the resistance R of a thermistor and its absolute temperature T was originally derived theoretically in the form $R = a\exp(b/T)$ from considerations of the free energy of an electron gas. This was shown to differ considerably from experiment and was replaced by the equation $\log R = \log A - C\log T + B/2{\cdot}303T$ by Becker, Green and Pearson (1947), who had noted that the graph of $\log R$ against $1/T$ increased linearly in slope with temperature. Bosson, Gutmann and Simmons (1950) have shown, however, that the resistance of a thermistor, at a given temperature, calculated from this formula, can differ by as much as 16·6% from the experimental value, and have suggested the substitution of $\log R = A + B/(T+\alpha)$, which gives a maximum discrepancy of about 2·5%.

In view of the uncertainties introduced by these formulae, the thermistors used were calibrated throughout the required temperature range against standard N.P.L. certificated thermometers, making use also of the ice point, the steam point and the transition point of sodium sulphate decahydrate, whose value (32·38°c) is accurately known. It being necessary to calibrate the thermistor in the same medium as it was to be used, it was placed, together with a stirrer and the standard thermometer, in a thin-walled glass tube, which was then partially immersed in an oil bath which was electrically heated and stirred. For measurements at the standard temperatures the thin-walled tube was placed in contact with the standard substance in a thermos vessel and left for a long period to attain temperature equilibrium.

Resistance measurements were made by having the thermistor as one arm of a precision Wheatstone bridge having, as nearly as possible, equal arms. The criteria of Callendar (1910) for maximum bridge sensitivity were observed. The thermistors used were of the Standard Telephones type F.2311/300 having a dissipation constant of 0·62 mw/°c in air. The bridge was energized from an accumulator through a potential divider arranged so that the power consumed by the thermistor did not exceed 4 μw and did not raise its temperature by more than about 0·006°c. To secure good sensitivity while using this very small power input to the bridge, a mirror galvanometer was used as a balance indicator.

At a temperature of 35°C, which was commonly used, this type of thermistor has a temperature coefficient of resistance of 0·03 per °C, which is eight times that of platinum. The heat sensitive bead was about 1 mm in diameter and, together with its glass mounting, weighed 1·8 g, having a thermal capacity of only 1·4 joules per °C. The thermistor and its leads were supported in a rigid holder so that the whole was free from stress. This is essential if results are to be reproducible. Two similar thermistors were used, one to measure the ambient temperature and the other the temperature at different points in the air jet. They were attached to the compound slide taken from a watchmaker's lathe, capable of horizontal motion in two directions at right angles. The slide could be raised or lowered on a stout steel rod fixed to a heavy base. The position of the thermistor in the air jet was determined by means of a cathetometer.

The experiment was first carried out in a small room having closely fitting blinds over the windows. Even on days of little wind, however, it was found that the temperature at a fixed point in the air jet fluctuated continually in a chance manner, the jet acting as a very sensitive draught detector. The source of heat and the thermistors were then placed in a cupboard, open on one side. A hole about 10 cm in diameter was placed in the roof of the cupboard vertically above the heat source, so that the hot air could escape without setting up a convective circulation in the observation chamber. This system reduced draughts, but they were still troublesome. It became obvious that they would have to be further reduced and the flow of hot air in the jet made visible. The extent of the influence of draughts would then be observed clearly, together with the efficacy of attempts to remove them. Further, the type of flow, whether laminar or turbulent, could be decided at once. The change in refractive index caused by heating the air was extremely small since the maximum heat output was only 20 watts. For the first trial a shadowgraph method was used, light from a motor-car headlamp bulb with a short filament, slightly over-run, passing through the air jet and striking a white screen. Such a system shows changes in illumination caused by the second derivative, $\partial^2\rho/\partial z^2$, of the density ρ with respect to distance z in the direction of flow. The method failed to delineate the path of this hot air stream at distances of more than about 20 cm from the weak heat source. A method of greater sensitivity was required.

At the request of Dr. O. G. Sutton to the Meteorological Research Committee, a high quality concave spherical mirror of 16 in. aperture and 200 in. radius of curvature was made available, and used to construct an efficient Schlieren system. The single mirror arrangement was selected for reasons of economy, in view of the large field of view desired.

The system used was similar to that of Taylor and Waldram (1933). A (figure 1) is a mercury high pressure lamp sending light through a condensing system consisting of two camera lenses B, C, of 9 cm aperture, with an iris diaphragm D placed between them. This light was focused to act as an intense source at E, the centre of curvature of the high grade spherical mirror FG. The reflected light could be blocked out at K, if desired, by the adjustable knife edge O. The distance from E to the mirror exceeded 500 cm. In figure 1 the separation of K from E, which was about 2 cm, has been exaggerated to allow the rays to be shown clearly. The hot air stream was produced in the centre of the region HJ. Since the rays traversed the hot air twice and were twice

deflected in the same direction, a magnification of deflection resulted. This principle was very useful to increase the minute angular deflections produced by the small amounts of heat used. After internal reflection by the prism L the light was received on the screen MN, made of ground glass or tracing cloth. This could be replaced by a camera when desired.

Consider the case of a medium of uniform density in front of the mirror. As the knife edge O was moved forward to obscure more of the image at K, a uniform darkening of the screen was caused, provided that O was at exactly the same distance from the screen as K, since every point of K received light from every point of the mirror. If, however, the space HJ was of uniform density except for a small region having a density gradient, this would be accompanied by a gradient of refractive index and the small region would act like a prism. Light passing through it would be deviated towards or away from the knife edge, causing the screen image of this small region to appear darker or brighter than the surrounding area of constant density. With the knife edge vertical, the horizontal components of density gradients could be

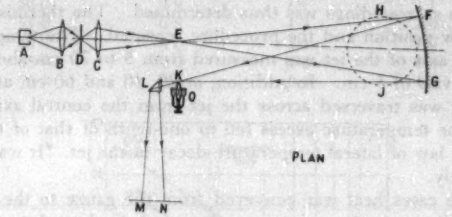


Figure 1.

detected most readily. A density gradient which caused light to be deviated along the knife edge caused no change in the illumination of the screen and so could not be observed.

Before the heated gauze, source of the hot air jet, was switched on, the knife edge was adjusted to allow a faint uniform illumination over the image of the mirror in the screen. In this faint light the thermistors could be seen outlined by diffraction bands. When the air jet was started, the thermistor could be seen in it and the type of air flow observed. The thermistor had initially been placed vertically above the centre of the heating gauze by using a plumb line. The Schlieren method showed whether the sensitive bead was truly central in the air stream and when the latter was disturbed by the slightest draught.

The Schlieren mirror and the heating gauze were now placed in a large wooden box, approximately 1 m × 1 m × 2 m tall. The heated air passed out of a hole in the top, whilst cool air entered through a large opening in the bottom, passing through two screens of perforated zinc to remove eddies before arriving at the gauze. The front of the box had two half doors, one above the other, to give access. The upper door had a large aperture covered by a sheet of selected plate glass. With this door open, light could fall directly on the mirror and adjustments could be made inside the box. With the door closed to exclude draughts, the image on the screen showed dark horizontal bars about 2 cm wide,

with clear spaces about 3 cm wide between them. These were due to regular changes in refractive index caused by the rollers used in the manufacture of the plate glass. It was easy to arrange that the important region round the thermistor appeared in a clear space. Further, the dark bars could be lightened by reducing the Schlieren sensitivity. On windless nights it was often possible to work with this door open. By contrast, at times of strong winds, with the door closed, and with closely fitting blinds over the windows, the air jet would be too disturbed by draughts and mechanical vibration to make dependable measurements. On hot, sunny days, even though windless, one wall of the laboratory would become warm, causing a disturbing circulation of the air in the room. All these disturbances were made visible by the Schlieren method.

The measurement procedure was to place the jet thermistor at the required height above the gauze, as measured by the cathetometer, and to check its position vertically above the centre of the gauze by means of the plumb line. The air stream was then observed visually by the Schlieren method, and conditions of steady flow awaited. When these were attained, the jet and ambient temperatures were measured, using their respective thermistors. The temperature excess of the jet over its surroundings was thus determined. The thermistors were then raised to a new position and the procedure repeated. The temperature excess on the central axis of the jet was measured from 5 to 70 cm above the heating gauze, at intervals of 5 cm. In addition, at 20, 40 and 60 cm above the gauze the thermistor was traversed across the jet from the central axis to beyond a point where the temperature excess fell to one-tenth of that of the axis. This was to test the law of lateral temperature decay in the jet. It was done for one heat output only.

In all these cases heat was conveyed from the gauze to the thermistor by convection, radiation and conduction. It was desired to determine the effect of heat transfer by convection alone. This was done at the different distances from gauze to thermistor where the temperatures due to the combined processes of heat transfer had already been measured. The method used was to have exactly the same conditions as before, except that the thermistor was placed at the same distance below the gauze as it had previously been above. In this way the temperature excess due to radiation and conduction was assessed, and by subtraction the effect of convection alone was determined.

It may be asked why the horizontal heating gauze used was not made smaller than 3·5 cm square so as to make the theoretical extension to the case of a point source easier. This was because, with a smaller heater, the region of maximum temperature on the axis of the jet was very small, so that a transverse movement of the thermistor of only 1 mm might cause its readings to change by several degrees when near the gauze. In short, the location and measurement of the point of maximum temperature became more difficult, and increasingly liable to inaccuracy as the area of the gauze was decreased.

§ 3. Results

The results for the excess of temperature $\Delta\theta$ of the central axis of the jet above its surroundings at various heights z above the heating gauze are given in table 1 for three different heat outputs, 4·87, 9·92 and 20·37 w. They have been corrected for the effect of radiation and conduction, and represent the contribution of convection alone. The results are plotted in figure 2. The

ambient temperatures at which these results were taken were all between 10·4 and 16·0°c. In figure 2 both $\Delta\theta$ and z are plotted on logarithmic scales; since the resultant graph is curved, its equation is clearly not given by $\Delta\theta = c/z^n$, c being a constant. An attempt was therefore made to find a theoretical point

Table 1. Points on Central Axis of Jet

Heat output (w)	4·87	9·92	20·37
z (cm)	Excess temperature $\Delta\theta$ (°c)		
5·0	7·8	12·9	20·5
10·0	5·2	7·7	12·2
15·0	3·4	5·8	9·2
20·0	2·7	4·7	7·2
25·0	2·1	3·4	5·4
30·0	1·6	2·8	4·3
35·0	1·4	2·3	3·8
40·0	1·3	2·1	3·1
45·0	1·1	1·7	2·6
50·0	0·9	1·4	2·1
55·0	0·8	1·3	2·0
60·0	0·8	1·1	1·7
65·0	0·7	1·0	1·4
70·0	0·6	0·9	1·2

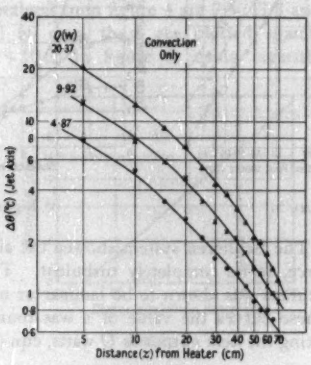


Figure 2.

source, at a distance a below the heating gauze, from which the airstream could be assumed to rise conically. If this was correct the $(\Delta\theta, z)$ relation would be

$$\Delta\theta = c/(z+a)^n. \qquad \ldots\ldots(1)$$

An approximate value of a was found by selecting three typical points on each of the curves of figure 2 in turn. For one curve let the coordinates of the three points be $(\Delta\theta_1, z_1)$, $(\Delta\theta_2, z_2)$ and $(\Delta\theta_3, z_3)$. Then

$$\frac{\log\Delta\theta_1 - \log\Delta\theta_2}{\log\Delta\theta_1 - \log\Delta\theta_3} = \frac{\log(z_2+a) - \log(z_1+a)}{\log(z_3+a) - \log(z_1+a)}.$$

Let the first fraction be called P and the second S. Inserting the numerical results for $\Delta\theta$, the value of P could be found. A number of trial values of a were inserted in S and so numerical values of S were calculated, care being taken to use only those values of a which gave small values (positive or negative) of $S-P$. The graph of $S-P$ against a was found to be linear and the required value of a which made $S-P=0$ was read from it. This was done for each of the curves of figure 2, and the average of the three values of a found to be 10·2 cm.

Still retaining the same meaning for z, the distance of any point on the axis of the jet from the theoretical point source was then $(z+10{\cdot}2)$ cm. A new graph of $\Delta\theta$ against $z+10{\cdot}2$ was then drawn (figure 3). The three resultant curves are seen to be linear, except for the last five points at 20·37 w and the

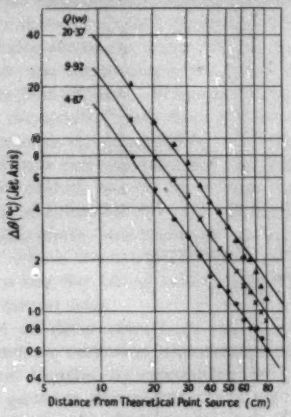


Figure 3.

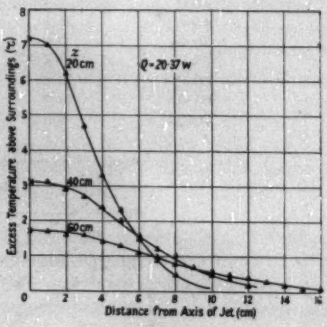


Figure 4.

last three at 9·92 w. The Schlieren system showed the air flow at these points, distant from the source, to be completely turbulent. The flow for all points falling on the linear curves was shown to be laminar or only slightly turbulent. From the slope of these curves the value of n was found to be 1·52 (to two decimal places). Writing the heat output as Q watts, eqn (1) was then re-written as

$$\Delta\theta = kQ^h/(z+10{\cdot}2)^{1{\cdot}52} \qquad \ldots\ldots(2)$$

k and h being constants.

It was desired to find the relation between the temperature on the central axis and the mean temperature across the jet at the same height. Temperatures across the jet at various distances x were measured by traversing the thermistor by means of the compound slide to a point beyond that at which the temperature fell to one-tenth of the axis value. Results, taken at intervals of 1 cm for a heat output of 20·37 w, are given in table 2 and plotted in figure 4.

The central axis excess temperatures, $\Delta\theta$, for these heights above the heater are given in table 1. From figure 4 the area under each curve, up to the point whose excess temperature was $\Delta\theta/10$, was calculated by Simpson's rule. The radius r of the jet, at each height explored, taken as the horizontal distance of this

point from the axis, was also measured from the curve. Dividing the area by r, the mean excess temperature $\overline{\Delta\theta}$ across the jet was determined for each height. The mean of the results gave $\Delta\theta = 1{\cdot}80\overline{\Delta\theta}$, suggesting a transverse temperature distribution in the jet similar to the normal law of error.

It was now possible to convert the large number of axial excess temperatures $\Delta\theta$, already found experimentally, into mean excess temperatures $\overline{\Delta\theta}$ across the jet. The equation giving the variation of $\Delta\theta$ with height above the theoretical point source also held for $\overline{\Delta\theta}$ if the constant k was changed to p, where $k = 1{\cdot}80p$.

Table 2. Transverse Excess Temperatures across the Jet due to Convection only

x (cm)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
z (cm)	Excess temperature above surroundings (°c)															
20	7·0	6·2	4·7	3·3	2·3	1·5	0·9	0·5								
40	3·1	2·9	2·7	2·4	2·0	1·6	1·2	1·0	0·7	0·5	0·4	0·2				
60	1·7	1·6	1·6	1·4	1·3	1·1	1·0	0·8	0·7	0·6	0·5	0·4	0·3	0·2	0·1	0·1

A series of curves was now drawn connecting $\Delta\theta$ and Q, the heat output in watts, on logarithmic scales. From the slopes of these curves, up to a height of 45·2 cm above the theoretical source, the average value of h in eqn (2) was found to be 0·675. The mean value of k, up to the same height, was 164·3, giving a corresponding value for p of 91·3.

The values of r obtained from figure 4 are 7·3, 11·3 and 15·0 cm for vertical distances 20, 40 and 60 cm; these are plotted against distance from the theoretical source $(z + 10{\cdot}2)$ in figure 5, using logarithmic scales for each

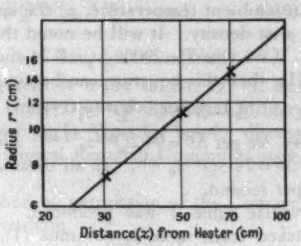


Figure 5.

variable. The curve is a straight line representing an equation of the type $r = A(z + 10{\cdot}2)^B$. From the slope of this curve $B = 0{\cdot}85$, and from the intercept on the horizontal axis A was calculated to be 0·40.

Hence, if we call the height of any point in the jet above the theoretical point source z_0 where $z_0 = z + 10{\cdot}2$, the final results of this series of experiments are contained in the equations

$$\overline{\Delta\theta} = 91{\cdot}3Q^{0{\cdot}675}z_0^{-1{\cdot}52} \qquad \text{and} \qquad r = 0{\cdot}40z_0^{0{\cdot}85}.$$

The value of z_0 in terms of z clearly applies only to the particular size and shape of heater used in these experiments, but it is believed that the other results are general.

§4. Comparison of these Results with Theory

This problem has been treated theoretically by Schmidt (1941) and Sutton (1948, 1950). Both assumed the mixing length concept in dealing with the formation of a vertical air stream, but whereas the first used infinite series in his solutions, the second obtained more direct results by assuming a close similarity between the spread of a hot gas stream and the diffusion of a cold gas in the atmosphere, provided the temperature of the hot gas was not much greater than its surroundings. He pointed out that in adapting the diffusion analogy to the case of the dispersion of a hot gas stream, the velocity of the natural convection, which decayed to zero with the excess temperature of the jet, could be considered as the wind involved in the diffusion. The reason for using only a small rise of temperature was so that the density of the gas could be considered constant. In the experiments just described, the maximum temperature rise was 20·5°c, and a typical ambient temperature was 13°c, so that the maximum density change was only 7%.

For the variation of the radius r of the jet with z_0 Schmidt assumed a relation $r = \text{const}\, z_0$, while Sutton calculated that $r = 1{\cdot}5Cz_0^{m/2}$ and, from the results of observations on smoke clouds in the atmosphere, he took $m = 1{\cdot}75$, so obtaining the result $r = 1{\cdot}5Cz_0^{0{\cdot}875}$, where C is a constant. If the present writer's experimental results for r and z_0 are inserted in the latter equation, the mean value obtained for C is 0·244. The writer found the power of z_0 to be 0·85. Using this result, the mean value for C is 0·269.

Sutton (1950) calculated a value for $\overline{\Delta\theta}$ which reduces to

$$\overline{\Delta\theta} = 5{\cdot}94 \times 10^{-3} T^{1/3} Q^{2/3} C^{-5/3} (c_p \rho)^{-2/3} z_0^{-1{\cdot}46}$$

where T is the absolute ambient temperature, c_p the specific heat of the air at constant pressure and ρ its density. It will be noted that $\overline{\Delta\theta}$ is not sensitive to the value of T, c_p or ρ. If we take $T = 290°\text{K}$, $c_p = 0{\cdot}24$ and $\rho = 1{\cdot}205 \times 10^{-3}$ g cm^{-3} at 20°c and 760 mm Hg, then this equation can be simplified to

$$\overline{\Delta\theta} = 8{\cdot}99 C^{-5/3} Q^{2/3} z_0^{-1{\cdot}46}.$$

If we take C to be 0·244, we get $\overline{\Delta\theta} = 94{\cdot}2 Q^{2/3} z_0^{-1{\cdot}46}$. Alternatively, if C is taken as 0·269, the result is $\overline{\Delta\theta} = 79{\cdot}5 Q^{2/3} z_0^{-1{\cdot}46}$. In all these equations Q is taken to be in watts, or joules per second.

To test these formulae table 3 was produced. In this are given the experimental results taken from table 1, columns (1), those calculated from $\overline{\Delta\theta} = 94{\cdot}2 Q^{2/3} z_0^{-1{\cdot}46}$, columns (2), and those calculated from $\overline{\Delta\theta} = 79{\cdot}5 Q^{2/3} z_0^{-1{\cdot}46}$,

Table 3. Comparison of Experimental and Calculated Results

Q (w)	4·87			9·92			20·37		
	Mean excess temperature $\overline{\Delta\theta}$ across jet								
z_0 (cm)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
15·2	4·3	5·1	4·3	7·2	8·2	6·9	11·4	13·2	11·2
20·2	2·9	3·4	2·8	4·3	5·4	4·6	6·8	8·7	7·4
25·2	1·9	2·4	2·0	3·2	3·9	3·3	5·1	6·3	5·3
30·2	1·5	1·9	1·6	2·7	3·0	2·5	4·0	4·9	4·2
35·2	1·2	1·5	1·3	1·9	2·4	2·0	3·0	3·9	3·3
45·2	0·8	1·0	0·9	1·3	1·7	1·4	2·1	2·7	2·3
50·2	0·7	0·9	0·8	1·2	1·4	1·2	1·7	2·3	1·9

columns (3). To make comparison easy, the experimental values of $\Delta\theta$ in table 1 have been divided by 1·8 to give $\overline{\Delta\theta}$ in table 3, and z_0 has been tabulated in place of z.

The values for $\overline{\Delta\theta}$ are accurate to the first decimal place in each column. The experimental results are closely fitted by the formula $\overline{\Delta\theta}=91{\cdot}3Q^{0{\cdot}675}z_0^{-1{\cdot}52}$ taken from them and described towards the end of § 3.

Clearly, the theoretical formula of Sutton used in columns (3), with C as 0·269, gives the closest approach to the experimental results. Comparison with Schmidt's formula $\overline{\Delta\theta}=\text{const}\,z_0^{-5/3}$ can only be made in terms of z_0, since he used only one, unstated, value of Q, and his figures for $\overline{\Delta\theta}$ were only proportional to the actual values, the constant of proportionality not being given. It is clear that on these grounds Sutton's formula agrees better with the experimental results than does Schmidt's. The experimental results give $-1{\cdot}52$ as the power of z_0, while Sutton and Schmidt calculate $-1{\cdot}46$ and $-1{\cdot}67$ respectively.

§ 5. Conclusions

1. Sutton's formulae

$$\overline{\Delta\theta}=5{\cdot}94\times10^{-3}T^{1/3}Q^{2/3}C^{-5/3}(c_p\rho)^{-2/3}z_0^{-1{\cdot}46} \qquad \text{and} \qquad r=1{\cdot}5Cz_0^{0{\cdot}875}$$

give very good agreement with experimental results when $C=0{\cdot}269$. With $T=290°\text{K}$, $c_p=0{\cdot}24$ and $\rho=1{\cdot}205\times10^{-3}\ \text{g cm}^{-3}$ these become

$$\overline{\Delta\theta}=79{\cdot}5Q^{2/3}z_0^{-1{\cdot}46} \qquad \text{and} \qquad r=0{\cdot}40z_0^{0{\cdot}875}$$

respectively.

2. The corresponding formulae derived directly from the experimental results are

$$\overline{\Delta\theta}=91{\cdot}3Q^{0{\cdot}675}z_0^{-1{\cdot}52} \qquad \text{and} \qquad r=0{\cdot}40z_0^{0{\cdot}85}.$$

Acknowledgments

Grateful acknowledgment is made to Dr. O. G. Sutton, Director of the Meteorological Service, lately Dean of the Royal Military College of Science, and to Associate Professor G. D. West for helpful discussions and for facilities provided. The Meteorological Research Committee is thanked for the loan of the Schlieren mirror.

This paper is published by permission of the Commandant, Royal Military College of Science, Shrivenham.

References

Becker, J. A., Green, C. B., and Pearson, G. L., 1947, *Bell Syst. Tech. J.*, **76**, 170.
Bosson, G., Gutmann, F., and Simmons, L. M., 1950, *J. Appl. Phys.*, **21**, 1267.
Callendar, H. L., 1910, *Proc. Phys. Soc.*, **22**, 220.
Schmidt, W., 1941, *Z. angew. Math. Mech.*, **21**, 265, 351.
Sutton, O. G., 1948, *Quart. J. R. Met. Soc.*, **74**, 13; 1950, *J. Met.*, **7**, 307.
Taylor, H. G., and Waldram, J. M., 1933, *J. Sci. Instrum.*, **10**, 378.

The Design and Operation of a Double Magnetic Lens Beta-Ray Spectrometer

By G. P. RUNDLE, J. ELLIS, T. C. GRIFFITH and H. S. TOMLINSON
Department of Physics, University College, London

*Communicated by H. S. W. Massey; MS. received 1st July 1953, and in amended form 28th September 1953. Read before the Physical Society at the Science Museum, London, 15th May 1953**

Abstract. Details are given of the design of a double magnetic lens spectrometer. Both coils of this spectrometer are capable of very fine adjustment in five degrees of freedom. Extensive work has been carried out on the testing of this instrument. This has included a study of the variation of spherical aberration with coil separation, the use of a quick and accurate experimental method of plotting trajectories, and an investigation of the extremely critical nature of coil alignment. Experimental confirmation of the predictions of Deutsch and his co-workers has also been obtained, and ring focus has been incorporated with notable success.

This spectrometer has a much lower spherical aberration than a short lens and gives a performance comparing not unfavourably with long lens types.

§ 1. Introduction

In recent years a variety of different types of beta- and gamma-ray spectrometers have been designed and built in many nuclear physics laboratories. Their use has, in the main, been confined to the study of beta- and gamma-ray spectra. Since spectrometers can be built with emphasis on a variety of different characteristics (viz. transmission, resolution, etc.), the design of a particular spectrometer must be carried out giving careful consideration to the nature of the experiment for which it is to be used.

The spectrometer described here was designed for the purpose of investigating the nuclear scattering of electrons and positrons, of energies up to 4 MeV, by thin metal foils. In such an experiment an intense beam of nearly monoenergetic particles is required. It is also necessary for the particles to be incident on the scatterer in directions which can be accurately determined. A magnetic lens spectrometer of the type designed by Deutsch *et al.* (1944) has high transmission at moderate resolution, and it is possible to place the scattering foils so as to intercept all the transmitted beam in a region essentially free of the magnetic field of the coil. The trajectories of the particles will, therefore, be essentially linear and their directions, relative to the foil, readily calculated from the geometry. This idea of using a lens spectrometer for scattering experiments was first used by Bothe and Ratzel (1940); they used a single, thin, coil and studied the scattering of electrons of energies up to 2 MeV.

The resolution of a lens spectrometer is improved by reducing the spherical aberration. This can be achieved, without loss of transmission, either with a long coil similar to the one described by Siegbahn (1944) or with two or more

* For a report of the symposium see Klemperer (1953).

separated coils as were used by Quade and Halliday (1948). Further, coils free of iron have the dual advantage that the spherical aberration is lower than for iron-clad coils, and that the relationship between the focused momentum and the current in the windings is linear. These considerations led to the decision that the best type of lens for the scattering experiment would be one with two identical coils, free of iron, each about the same width as the single thin lens of Deutsch *et al.*

For high intensity in the scattering experiment a large initial angular range may be accepted giving high transmission at moderate resolution. In other experiments a small initial angular range can be selected, thus enabling the instrument to be used as a high-resolution spectrometer. The resolution and transmission of the instrument compare favourably with all other spectrometers in use except the double focusing type which, in any case, would be unsuitable for the scattering experiment.

The two coils can be accurately aligned in space to within 0·5 minute of arc in any direction, and may be separated by any distance from 0 to 10 in. These fine adjustments enabled a detailed study to be made of the effect of coil alignment and separation on the resolution and transmission of this type of spectrometer. Ring focus has also been incorporated with notable success in improving the resolution.

The spectroscopic nomenclature of Persico and Geoffrion (1950) has been adopted throughout this paper.

§ 2. Design and Construction of the Coils

2.1. *Design*

The design of each coil was based on calculations which assumed them to be placed in contact as a single unit; an allowance was then made for an increase of about 25% in the focusing current, this being necessary due to a reduction in focusing efficiency as the coils are separated. The dimensions of the combined coil were so chosen that the current density in the windings would not exceed 1500 A in^{-2} when 4 MeV electrons are focused under the condition $u=v=2f=50$ cm. This calculation was carried out by determining the field H_z along the axis using the formula given by Deutsch *et al.* (1944), estimating $\int H_z^2\,dz$ graphically, and testing whether this satisfied the formula $f=4(H\rho)^2/\int H_z^2\,dz$.

The coils are identical, and each is wound in eight sections. All sections consist of eight layers of asbestos-covered copper wire of rectangular section, 0·142 in. by 0·071 in. Between each section open ducts $\frac{1}{8}$ in. wide have been allowed for cooling with a draught of forced air (figure 1). The wire is wound on a central brass former $\frac{1}{4}$ in. thick, with no support for the faces of the coils other than paper insulation separating the layers. This allows the coils to be pushed together so that their faces are in contact. The coils are wound as three sections, normally connected in series, so that if desired the outer sections only may be employed. The section ends are brought out on the outer face of each coil. Extra insulation and rigidity is obtained by immersing the coils in shellac and baking them.

Each completed coil has a weight of 500 lb, a width of 5·5 in., and 2057 turns of wire. The internal and external diameters of the windings are $9\frac{1}{4}$ in. and 24 in. respectively. The maximum allowed current density is 1500 A in^{-2} with forced

cooling and 1100 A in^{-2} without cooling. The cold resistance of the combined coil is 15·4 ohms and its inductance (at 50 c/s) is 5 henrys. The maximum power dissipation in the coils is 3·5 kw.

Two spun $\frac{1}{16}$ in. thick aluminium covers are mounted on the outer face of each coil. There are two inlets, 2 in. in diameter, in each cover through which a forced air draught from four blowers may be delivered. No point in the windings is further than $\frac{1}{4}$ in. from the air draught, and the maximum temperature rise inside does not exceed 50°C.

The current through the coils is supplied from a motor-generator set, capable of delivering 6 kw at 300 v. A field current of 0·67 A at 300 v is supplied from an exciter coupled to the same shaft as the generator. The exciter field requires 200 mA maximum excitation which is obtained from three 6A3 power valves of a d.c. amplifier. The current is automatically stabilized by means of a double loop feedback system similar to the one described by Danforth (1939). Voltage stabilization is obtained by applying the difference between the generator output voltage and a reference voltage to the grid of the first valve of the current amplifier which, in turn, controls the plate current of the power valves supplying the exciter field. An additional reference voltage, in series with the first, derived from a conventional standard resistance, potentiometer, galvanometer and photocell system, constitutes the current stabilizing loop. A current stable to better than one part in 2000 is obtained in this way. The coil current is measured by means of a standard manganin resistance in series with the coils, and a Tinsley potentiometer.

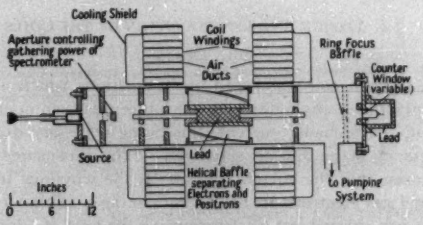


Figure 1. The spectrometer.

As a safeguard against the large back e.m.f. which would develop if the coil current were suddenly broken a 300 v 2 A selenium metal rectifier is permanently connected across the coil with its polarity so arranged that under normal operation it is not conducting.

The spectrometer chamber is constructed from $\frac{1}{4}$ in. thick aluminium sheet. The chamber and flanges are machined as one unit. By means of spigots on the flanges and registers on the end plates both the Geiger–Müller counter and the source can be placed centrally on the axis of the chamber. The length of the chamber is 40·5 in., wall thickness after machining $\frac{1}{8}$ in., and its inside diameter is 7·75 in. to within 0·005 in. along the entire length of the tube. The design of the baffles, which slide smoothly into the chamber, was based on a knowledge of the shape of the trajectories of the particles, found in the manner described in § 3.3. The helical baffle is $8\frac{3}{4}$ in. long, and houses a 6 in. long cylinder of lead for absorbing gamma-rays. There are six vanes each $\frac{1}{4}$ in. thick, the rotation along their length

being $35\frac{1}{2}°$. To reduce the scattering from material within the spectrometer tube all the baffles are constructed of aluminium. A pressure of 2×10^{-5} mm Hg can be maintained in the chamber by means of a two-stage oil diffusion pump backed by a mechanical pump.

The axis of the spectrometer is made to coincide with the direction of the horizontal component of the Earth's field, and neutralization of the vertical component is achieved with a pair of coils of design based on the calculations of Lyddane and Ruark (1939). These coils have bent back ends and produce a field uniform to within 1% over the region occupied by the spectrometer chamber.

2.2. *Coil Adjustment and Support Mechanism*

The construction of the coil mounting assembly allows each coil five degrees of freedom with respect to the fixed spectrometer tube. These movements are finely adjustable. Two of these adjustments are horizontal and vertical displacements in a plane perpendicular to the spectrometer axis, and two are rotations about horizontal and vertical axes in this plane. A further movement allows a separation of the coils along a line parallel to the axis of the spectrometer. The half cross section through the centre of one of the coil supports (figure 2) shows the main design features. The construction material is aluminium alloy and brass.

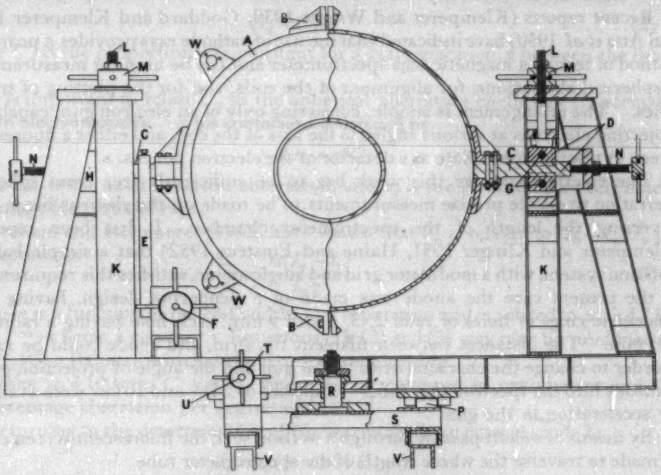


Figure 2. The coil mounting system.

The coils are held by the semi-circular bands A, clamped by the brackets B. Double brackets riveted to the bands carry the collinear coil axles C, which are housed in the split bearings D. Controlled rotation of the coils about the axis of these axles, through $\pm 6°$ from the perpendicular to the spectrometer tube axis, is obtained by movement of the coupled lever arm E, the moving end of which is controlled by the threaded guide F. Pin bearings G join the axle bearings to their

housing boxes H, the latter being sliding fits in the cradle sides K. Each bearing box is suspended as a unit from the threaded rods L, which hang from the top of the cradle sides. The threaded capstan wheel M and its lock nut allow the coil axles to be raised or lowered independently through the range ± 1 in. about the spectrometer tube axis with absolute rigidity and without axle jamming. The axles are moved through their bearings parallel to their axis by the threaded rods N. The range of movement is ± 1 in. about the axis of the spectrometer tube. Each cradle rotates on eight roller supports P about the centre bearing R, on the cross member of the long wheel base support trolley S. The threaded capstan wheels, T, which operate the screwed rods U set on either side of the centre bearing, control this rotation over the range of 12°. The support trolleys are telescopic to allow the inner coil faces to meet, and run on the machined brass rails V, which provide a maximum spacing of 10 in. between the inner coil faces. These rails, which are accurately parallel to each other, and to the spectrometer tube axis, are bolted to the top of the main trolley carrying the whole spectrometer. Four coil spacing rods, provided with universal clamps, fit into sockets W and oppose the electromagnetic attraction of the coils.

§3. Investigation of the Characteristics of the Spectrometer

3.1. *Cathode-Ray Technique*

Recent reports (Klemperer and Wright 1939, Goddard and Klemperer 1944, Van Atta *et al.* 1950) have indicated that the use of cathode rays provides a powerful method of testing a magnetic lens spectrometer and can be used for measurements of spherical aberrations, for alignment of the coils, and for the plotting of trajectories. The arrangement is simple, consisting only of an electron gun, capable of projecting electrons at various angles to the axis of the coil, and either a fluorescent screen or photographic plate as a detector of the electron pencils.

The electron gun for this work has to be sufficiently free from spherical aberration to enable precise measurements to be made on the electron beam after traversing the length of the spectrometer chamber. It has been reported (Klemperer and Klinger 1951, Haine and Einstein 1952) that a simple hairpin emission system, with a modulator grid and single anode, satisfies this requirement. In the present case the anode was made of a pepperpot design, having four concentric rings of holes of radii 2, 3, 6 and 9 mm, each hole having a radius of 0·05 mm. The distances between filament tip, grid, and anode could be varied in order to change the characteristics of the gun and the angle of projection of the electrons into the spectrometer tube. Supplies of 2 kev and 10 kev were available for acceleration in the gun.

By means of a shaft passing through a Wilson seal, the fluorescent screen could be made to traverse the whole length of the spectrometer tube.

3.2. *Measurement of Spherical Aberration*

Quade and Halliday (1948) have previously investigated the spherical aberration of a separated coil system for various separations of the lens coils, and found that this had a minimum value for a particular spacing. It was therefore decided to re-investigate, experimentally, the variation of the spherical aberration with spearation, in order to determine the appropriate optimum spacing for the spectrometer described here.

The gun filament was arranged with its tip in the position normally occupied by the source, and the moving screen replaced by one fixed in the plane usually occupied by the window of the Geiger–Müller counter. The coil current was obtained from large capacity accumulators and determined from the potential developed across a standard resistance, measured on a Tinsley vernier potentiometer. With this experimental arrangement the current necessary to focus successive rings of electron pencils was recorded for various coil separations.

Since the initial angles of the electrons are not given directly by the geometry of the gun, the moving screen was set at various distances from the gun, and the positions of the electron spots relative to the axis were measured with a cathetometer. This gave the initial angles α of the electron pencils with respect to the axis directly. Also, to determine whether any spurious effect might be present due to the focusing field, the initial angles were redetermined with and without the field for a number of coil separations. No variation in α was observed.

For quantitative measurement a spherical aberration coefficient A_α similar to that used by Quade and Halliday was defined as

$$A_\alpha = 100(I_0 - I_\alpha)/I_0 \,\% \qquad \ldots\ldots(1)$$

where I_α is the current required to focus rays making an initial angle α with the axis, and I_0 is the current that would be required to focus paraxial rays.

Since A_α is proportional to α^2 (Persico and Geoffrion 1950) we can write

$$A_\alpha = C_d \alpha^2 \qquad \ldots\ldots(2)$$

where C_d is the proportionality constant for a coil separation d between the inner faces.

It is important to relate C_d to the spherical aberration coefficient C_s normally used in electron optics. This coefficient is defined by

$$-\Delta z_f = \alpha^2 C_s$$

where Δz_f is the axial spherical aberration, and z_f is the distance between object and image for paraxial rays.

Graham and Klemperer (1952) have shown that C_s is related to $(I_0 - I_\alpha)/I_0$ by the expression

$$\frac{C_s}{z_f} = \frac{m(I_0 - I_\alpha)}{I_0 \alpha^2} \qquad \ldots\ldots(3)$$

where m is a function of the coil widths and separation and is related to z_f and I by $z_f = k I^{-m}$, k being a constant. Equations (1), (2) and (3) may now be combined to give $C_d/C_s = 100/mz_f$, which is the required relationship.

When $z_f = 100$ cm, $C_s = mC_d$, where C_s is measured in centimetres and C_d in percentage aberration per degree squared.

Returning to the determination of C_d, we obtain from eqns (1) and (2)

$$I_\alpha = I_0(1 - 0{\cdot}01\, C_d \alpha^2).$$

Thus C_s can be determined from a graph of I_α against α^2. Such a graph was drawn for various values of d for the complete coils and also for the outer sections only. The results are shown in figure 3.

The curves obtained have the same general shape as those of Quade and Halliday for small coil separation, but there is no evidence of the minimum reported by these authors. The explanation of this is to be found in the difference between the methods employed for defining the initial angles α of the electrons considered in the two cases. In the present experiment electrons at fixed initial angles, defined

in a region where the trajectories are essentially linear, were considered, the angles being the same for all coil separations. Quade and Halliday, however, considered electrons defined by fixed apertures (of different diameters) in a central baffle and, therefore, it is the lens field that determines the initial angles for each coil separation. As the coils are separated the field distribution along the axis broadens, the electron trajectories became flatter (§3.3) and therefore the initial angles increase. With increasing initial angles an increase of spherical aberration is to be expected, and eventually this increase outweights the decrease of spherical aberration that is to be expected at large coil separation (see below), and a minimum is therefore obtained.

A comparison of the spherical aberration of various spectrometers is shown in figure 4. The graph given by Deutsch *et al.*, from which curve A was calculated, showed the variation of spherical aberration with a parameter R. This parameter

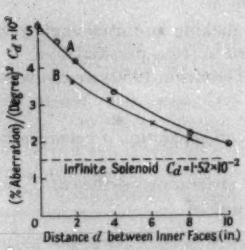


Figure 3. Variation of spherical aberration with separation. A, current in all windings; B, current in outer section only.

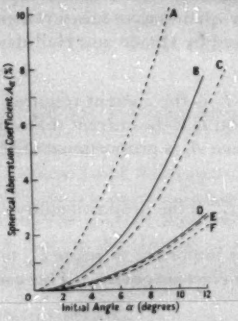


Figure 4. Comparison of the spherical aberration for various lens arrangements. A, thin coil of Deutsch *et al.*; B, zero separation ($d=0$); C, system of Quade and Halliday (optimum separation); D, $d=10''$; E, long lens of Graham and Klemperer; F, infinite solenoid.
– – – – previous lens systems.
——— present lens system for current in all windings.

corresponds to the off-axis distance of particles in the central plane of the lens, for a given initial angle, when the coil current is such that the particles would be focused on the counter window. Another parameter R' was defined as the off-axis distance in the same plane and for the same initial angle, but with zero current in the coils. Their ratio of R' to R was equal to 1·17, and hence their original curve was converted to that of figure 4 curve A by use of the relationship

$$\alpha = \tan^{-1}(R'/2f) = \tan^{-1}(0{\cdot}0234R)$$

Figure 4 curve C was also obtained from a graph of spherical aberration plotted against R, but in this case no indication of R'/R was given by the authors. An estimate was therefore made by comparing the different values of R'/R for various spectrometers with the overall length of the coil systems. Use was made here of

the figures appropriate to the thin lens of Deutsch *et al.* (1944), the long lens of Graham and Klemperer (1952), an infinite solenoid ($R'/R=\pi/2$), and a number of separations of the system being described. This indicated that for the Quade and Halliday (1948) system $R'/R\sim 1{\cdot}2$.

Qualitatively, the mechanism by which the separation of the coils reduces the spherical aberration (figure 3) may be explained in the following manner. The spherical aberration in a magnetic lens can be considered as arising from two distinct effects: (i) the 'cosine error' due to a variation of axial velocity with angle of projection; this results in trajectories of large initial angle crossing the axis nearer the source than paraxial paths; this error is defined as positive; (ii) the error due to a finite value of $\partial H_z/\partial r$, which we shall call the 'field error'; if this is positive, marginal rays pass through a more intense field, which results in stronger focusing and hence an increased aberration; however, if $\partial H_z/\partial r$ is negative the spherical aberration will be reduced (this differential is zero in the case of an infinite solenoid and we have only the cosine error).

Now it was shown by Siegbahn (1946) that

$$H_z(r, z)=H(z)-\tfrac{1}{4}r^2H''(z)+\ldots.$$

so that $\partial H_z(r, z)/\partial r\sim-\tfrac{1}{2}rH''(z)$; hence for a reduction of spherical aberration we require $H''(z)$ to be positive. On applying this to the case of the double lens system, it is readily seen that as the coils are separated the range of positive $H''(z)$ increases, thereby reducing the total spherical aberration.

The observed variation of C_d with d is in agreement with both the experimental work of Lauritsen and Christy (1948) and the theoretical work of Snowdon (1951). It might also be added that as the coils are separated the field distribution tends to that predicted by Glaser (1940) for zero spherical aberration; it has been shown since, however, that such a distribution would not act as a lens.

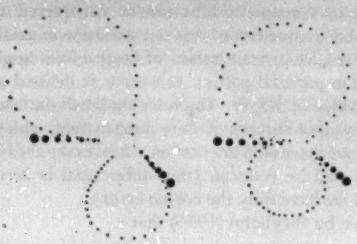
3.3. *Determination of Trajectory Shapes*

It was decided that the simplest and most direct method of determining the trajectories would be experimentally rather than analytically. The fluorescent screen was replaced by a photographic plate and, with the coil current set to focus electrons at fixed incident angles, exposures were made at various positions along the tube. For these exposures a shutter operated through a Wilson seal could be introduced in front of the gun as the exposure at each position of the film was completed. An exposure every 3 cm along the axis was found to be adequate. Using Ilford industrial x-ray film type B exposures of a few seconds sufficed.

The trajectory trace obtained in this way is a projection of the helical path on a plane perpendicular to the axis of the coil. The rotation can be determined at once from such a trace and, by measuring the distance of each exposed point of the photograph from the axis, the trace can rapidly be converted to a plot of the trajectory in a meridional plane rotating with the electrons. Two typical exposures are shown in figure 5: in each case two trajectories, of different initial angles, have been determined simultaneously.

It will be observed that the rotation is much greater for the spaced coil system than for a single coil, and that the trajectory is also much flatter; this latter point is more clearly shown by figure 6. The large black spots are caused by light from the filament coming through the holes in the anode cap with the electrons. Near the gun end of the tube these give more intense spots than the electrons.

Figure 6 shows some trajectory plots obtained from the films. In each case the gun was set to deliver two beams at approximately 7° and 10° to the axis, both beams being traced on the same film. In each case the coil current was that required to make the outer beam cross the axis at the normal point of focus. The effect of a thick coil, and of two spaced coils, in reducing the spherical aberration below that of a short coil is very clearly illustrated.



(*a*) (*b*)

Figure 5. (*a*) Typical trajectory traces for a single coil, (*b*) two coils separated by 10 in. between their inner faces.

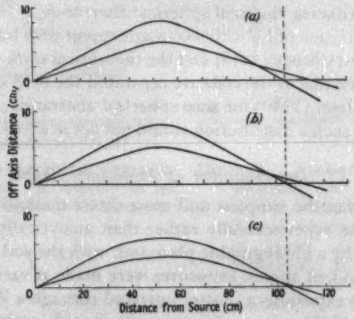


Figure 6. Trajectories in a plane rotating with the electrons for (*a*) a single coil, (*b*) two coils in contact, and (*c*) two coils with 10 in. between inner faces.

The great merit of this method is that a large number of trajectories can be recorded simultaneously, provided due attention is paid to possible overlapping of the small spots by the larger ones. It is also applicable to tracing initially off-axis trajectories since all that has to be done is to move the gun along the axis until the desired electron ray crosses the source plane at the desired 'off-axis' distance. With a suitable gun, initially skew rays could also be simply plotted.

The spectrometer baffles (limiting, scraping and helical) were designed and positioned from a knowledge of the trajectories gained in this way.

3.4. *Alignment*

With the spectrometer arranged with its axis parallel to the horizontal component of the Earth's magnetic field, the magnetic axis of each lens coil was

made coincident with the geometric axis of the spectrometer chamber by the cathode-ray technique described by Van Atta *et al.* (1950). This not only gives the correct alignment but also gives the current necessary in the degaussing coils for correct compensation.

With this initial adjustment the gun and screen were replaced by a 1 mm diameter source of thorium activated deposit and a Geiger–Müller tube of 1·5 mm diameter window, respectively. Final small improvements were then sought in the shape and peak height of the F line.

The extremely critical nature of the adjustment was immediately evident. Typical of the cases observed was the line shape with a maximum count at peak of 135 per minute which had poor resolution, and a second peak, nearly as high as the first in an adjacent position; a rotation of 15 minutes of arc of one coil increased the counting rate at peak to about 10 000 counts per minute, and improved the resolution from 6% to about 1·5%. Another example was an increase in counting rate of 80% for a rotation of only 5 minutes of arc. It was found that rotation of the coils had a much greater effect on the focused spot than a lateral displacement. The latter only caused a movement of the focused spot of the same order as the movement of the coil. This very critical nature of alignment has already been reported by Schmidt (1952).

It should be pointed out that, due to the possibility of the existence of compensating errors between the coils, each coil had to be aligned separately in order to obtain alignment for both directions of the lens current. This is essential when separating electrons and positrons by means of a helical baffle and current reversal.

3.5. *Calibration*

Since there is no iron in or near the coils, the current I required to focus an electron of given momentum $H\rho$ is proportional to the momentum. It is only necessary then to determine the current required to focus one conversion line, for which the $H\rho$ value is accurately known, in order to find the calibration constant k, defined by $k = H\rho/I$. However, as a check, k was determined using both the F and I lines of Th(B+C), and also the K line of ^{137}Cs. The calibration constant was found to be 967 G cm A^{-1} for the coils in their normal position (10 in. between inside faces). For the two coils together $k = 1355$ G cm A^{-1}, and for a single coil $k = 755$ G cm A^{-1}. In all three cases the coils were placed symmetrically between source and counter.

3.6. *Measurement of Magnetic Field Distribution*

The field distributions along the axis of the coils are shown graphically in figure 7 for various coil arrangements. The field distribution for the two coils together is compared with that calculated from the dimensions of the coil, using the formula given by Deutsch *et al.* (1944). The distributions have been normalized in such a manner that they represent the field strengths required to focus electrons of a given energy.

3.7. *Performance*

The present initial angular range ($7° < \alpha < 10°$) gives a gathering power of 0·39%. Should it become necessary, the 'flat' trajectories of the spaced arrangement will allow this to be greatly increased (cf. figures 1 and 6).

It was shown by Deutsch *et al.* (1944) that for a thin coil, free from spherical aberration, the full spread at half height due to geometry is given by

$$\Delta p/p = 0{\cdot}2(ms+w)/d'(m+1)$$

for optimum operating conditions, i.e. for $w/ms = 1$, where s is the source diameter, w is the window diameter, m is the magnification (1 in this case), and d' is the minimum value of R' (see §3.2) for the transmitted beam. For 10in. spacing $d' = 5{\cdot}5$ cm, hence $\Delta p/p = (s+w)/5{\cdot}5\,\%$, if s and w are expressed in millimetres.

The spread, which is experimentally determined, consists of the spread due to spherical aberration, which is independent of $s+w$, and the geometrical spread which is directly proportional to $s+w$. Thus a plot of experimental spread against $s+w$ should be linear and parallel to the plot of geometrical spread. Also, extrapolation to $s+w=0$ should give the contribution to the spread from spherical aberration only.

Using sources activated in the presence of 2 mc of radiothorium, we plotted a number of line shapes were plotted for various combinations of s and w. In figure 8 the spreads of those for which $w/ms = 1$ are plotted against $s+w$; so also are the spreads to be expected from geometrical considerations only.

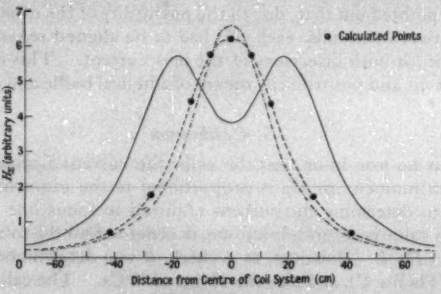


Figure 7. Distribution of axial field for various coil arrangements.
—·—·— single coil, — — — two coils together, ———— two coils with 10 in. spacing between inner faces. *Note.* The calculated points are for the two coils together.

It can be seen that the experimental curve and the geometrical curve are very nearly parallel, and that the contribution from spherical aberration is about 0·8%, which is in good agreement with the value found by the cathode-ray technique for the same angular range. Further, when $w \neq s$ the spread was increased compared with that corresponding to a similar value of $s+w$ in which s and w were equal. This gives direct confirmation of the conclusion arrived at by Deutsch *et al.* (1944) that for optimum performance the counter window w should be made of the same size as the image ($=ms$), or slightly larger if there is much spherical aberration.

Comparison with other spectrometers is difficult because often the gathering power, source size, etc., are not strictly comparable. One might, however, be permitted to assume that the 'typical high resolution' curve of Deutsch *et al.* (1944) represents the best resolution that could be obtained with their short lens. This has a spread of 2% for $s = 4$ mm, $w = 6$ mm and was obtained with the restricted angular range ($8° 20' < \alpha < 9° 40'$) and using the outer segments only of

the coils, two conditions which favour good resolving power. The spread of a similar curve (figure 9) (taken with the instrument described here) for an angular range of 3°, and using all the windings when the coils are separated by 10 in., is 1·3% when $s = 1$ mm and $w = 1{\cdot}5$ mm. The ratio of peak height to the continuous spectrum height on the right-hand edge is 15 : 1 in this case and about 8 : 1 for the curve of Deutsch *et al.*

Not unfavourable comparison can also be made with the long lens of Graham (1949), the spherical aberration of which is illustrated in figure 4. This long lens gives a spread of 1·47% for a gathering power of 0·5%, when $s = 0{\cdot}7$ mm and $w = 1{\cdot}5$ mm.

To summarize, it can be said that the spectrometer described in this paper has a much lower spherical aberration than a short lens, and gives a performance comparing favourably with long lens types.

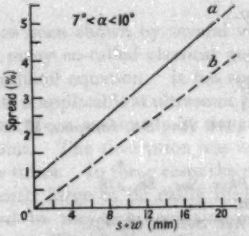


Figure 8. Variation of spread with $(s+w)$: *a*, experimental curve, *b*, spread to be expected from geometrical considerations only.

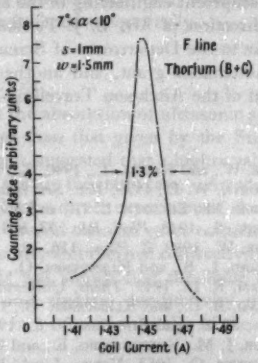


Figure 9. A typical experimental line shape with axial focusing.

3.8. *Ring Focus*

The discussion in § 3·7 referred only to cases in which momentum discrimination was achieved by means of a small circular axial aperture in the plane of the focus (known as axial focus). However, it has been shown by previous workers (Witcher 1941, Frankel 1948, Keller, Koenigsberg and Paskin 1950, Pratt, Boley and Nichols 1951) that as a result of spherical aberration a ring-shaped constriction is formed in the focused beam between the source and counter window. Further, if the exit aperture of the system is made in the form of an annular baffle placed about the position of this ring focus, the transmission can be increased for a given resolution.

The work in § 3.3 gave a good indication of the expected position of this constriction. The exact position was then found by exposing photographic plates at various positions along the axis. Using Ilford type G super sensitive x-ray film and sources irradiated by 30 mc of radiothorium, exposures of the order of 15 minutes were required to photograph the F line.

The reduction in the spherical aberration of the system obtained by separating the coils had the effect of bringing this ring constriction to within 6 cm of the position

of axial focus. An example of the improvement obtained is given below. With a source 1 mm in diameter and a ring focus aperture 1 mm wide a spread of 0·97% was obtained, compared with that of 1·58% when an axial aperture of 1·5 mm diameter was used. Also, the peak height in the ring focus case was 40% greater than with axial focusing. Thus an overall increase of 128% in the ratio of peak height to spread was obtained with ring focus, which is in agreement with the theoretical predictions of Keller, Koenigsberg and Paskin (1950).

Acknowledgments

The authors wish to thank Professor H. S. W. Massey and Dr. E. H. S. Burhop for the keen interest shown throughout the programme, Dr. C. Henderson for his help during its latter stages, and Dr. O. Klemperer of Imperial College, London, for his advice concerning the theory of aberrations. Further, we wish to acknowledge the competent engineering of the apparatus by the departmental workship under the direction of Mr. E. J. Faulkner. One of us (J.E.) wishes to express his thanks to the Department of Scientific and Industrial Research for the award of a maintenance grant, and another (G.P.R.) to Melbourne University for the award of the Aitchison Travelling Scholarship.

References

Bothe, W., and Ratzel, L., 1940, *Z. Phys.*, **115**, 497.
Danforth, W. E., 1939, *Rev. Sci. Instrum.*, **10**, 211.
Deutsch, M., Elliott, L. G., and Evans, R. D., 1944, *Rev. Sci. Instrum.*, **15**, 178.
Frankel, S., 1948, *Phys. Rev.*, **73**, 804.
Glaser, W., 1940, *Z. Phys.*, **116**, 19.
Goddard, L. S., and Klemperer, O., 1944, *Proc. Phys. Soc.*, **56**, 378.
Graham, R. L., 1949, *Thesis*, University of London.
Graham, R. L., and Klemperer, O., 1952, *Proc. Phys. Soc.* B, **65**, 921.
Haine, M. E., and Einstein, P. A., 1952, *J. Appl. Phys.*, **3**, 40.
Keller, J. M., Koenigsberg, E., and Paskin, A., 1950, *Rev. Sci. Instrum.*, **21**, 713.
Klemperer, O., 1953, *Nature, Lond.*, **172**, 61.
Klemperer, O., and Klinger, Y., 1951, *Proc. Phys. Soc.* B, **64**, 231.
Klemperer, O., and Wright, W. D., 1939, *Proc. Phys. Soc.*, **51**, 296.
Lauritsen, T., and Christy, R. F., 1948, *Phys. Rev.*, **75**, 536.
Lyddane, R. H., and Ruark, A. E., 1939, *Rev. Sci. Instrum.*, **10**, 253.
Persico, E., and Geoffrion, C., 1950, *Rev. Sci. Instrum.*, **21**, 945.
Pratt, W. W., Boley, F. I., and Nichols, R. T., 1951, *Rev. Sci. Instrum.*, **22**, 92.
Quade, E. A., and Halliday, D., 1948, *Rev. Sci. Instrum.*, **19**, 234.
Schmidt, F. H., 1952, *Rev. Sci. Instrum.*, **23**, 361.
Siegbahn, K., 1944, *Ark. Mat. Astr. Fys.* A, **30**, 1; 1946, *Phil. Mag.*, **37**, 162.
Snowdon, S. C., 1951, *Rev. Sci. Instrum.*, **22**, 878.
Van Atta, C. M., Warshaw, S. D., Chen, J. J. L., and Taimuty, S. I., 1950, *Rev. Sci. Instrum.*, **21**, 985.
Witcher, C. M., 1941, *Phys. Rev.*, **60**, 32.

The Absorption of Sound in Carbon Dioxide contained in Narrow Tubes

By L. E. LAWLEY*
Department of Physics, King's College, Newcastle-upon-Tyne

Communicated by E. G. Richardson; *MS. received* 14*th September* 1953

Abstract. Measurements have been made of ultrasonic absorption in carbon dioxide contained in narrow tubes of the order of 1 mm diameter. Frequencies used were between 40 and 120 kc/s. Results showed that over a considerable range of frequency/pressure on either side of the relaxation peak of CO_2 the total absorption was given by the sum of the Helmholtz–Kirchhoff tube absorption and the thermal relaxation absorption of the gas. The tube absorption, however, was found to be a few per cent higher than that calculated theoretically.

§ 1. Introduction

It has been shown by several workers that the absorption of ultrasonic sound by many so-called classical gases is higher than that given by the Stokes–Kirchhoff equation. It has sometimes been suggested that a higher value of viscosity is applicable at ultrasonic frequencies than in the static case. An account has been given previously by the author (1952) of a direct experimental check of this point. The absorption was measured in several classical gases contained in narrow tubes. In these cases the effect of viscosity and thermal conductivity was sufficiently large to be easily measurable. It was shown that the value to be ascribed to the term containing viscosity and thermal conductivity in the Helmholtz–Kirchhoff equation for absorption in a tube was usually of the order of 5% above the normal static value.

The work described here is an extension of the previous investigation to the case of a non-classical gas. The aim of this work is (i) to measure the absorption due to viscosity and thermal conductivity when thermal relaxation processes take place at the same time in the gas, (ii) to show whether or not the viscous and relaxation mechanisms occur independently in the gas and (iii) to provide an independent check under different conditions of the work of Fricke (1940) who obtained the relaxation curve for CO_2 in free space.

It has been shown theoretically by Herzfield and Rice (1928) that in most practical cases the total absorption in a free gas is the sum of that due to heat conduction, viscosity and thermal relaxation. Experimental work on carbon dioxide (Fricke 1940, Van Itterbeek 1938, Zartman 1949) shows that for low values of frequency/pressure the absorption follows a theoretical thermal relaxation curve and at high values of f/p it is given approximately by the classical heat conduction–viscosity equation. However, in the region where the absorption due to each of these causes is of the same magnitude the total absorption is very small, hence if second order effects occur under these conditions they could not be observed.

In the present investigation the gas was contained in a narrow tube so that even at the peak of the thermal relaxation curve the viscous absorption due to the

* Now at The Polytechnic, Regent Street, London W.1.

tube was of the same order, thus any absorption term arising due to the combined effects of heat conduction, viscosity and thermal relaxation should be immediately apparent. The total absorption under such conditions will be due to (i) thermal relaxation, (ii) Helmholtz–Kirchhoff tube effect due to viscosity and thermal conductivity and (iii) Stokes–Kirchhoff free gas plane wave classical absorption. In the present work classical absorption is very small indeed. The effect of the tube is confined to an area very close to the tube walls, so that effects (i) and (ii) are likely to occur independently and will be simply additive. Thus we have as a theoretical equation for the total absorption μ per wavelength ($\mu = 2\alpha\lambda$):

$$\mu = K_T X^{1/2} + K_G X X_m/(X^2 + X_m^2) \qquad \ldots\ldots(1)$$

where

$$K_T = 2(\pi p)^{1/2}[\eta^{1/2} + (\gamma - 1)(K/C_p)^{1/2}]/rfp^{1/2} \qquad \ldots\ldots(2)$$

$$K_G = 2\pi \boldsymbol{R} C_i/[(\boldsymbol{R} + C_0)(\boldsymbol{R} + C_\infty)C_0 C_\infty]^{1/2} = 2\mu_m. \qquad \ldots\ldots(3)$$

f is frequency, p pressure, r tube radius, η viscosity, K thermal conductivity, $\boldsymbol{R}$ gas constant, C_0 and C_∞ specific heats (at constant volume) at zero and infinite frequencies respectively, C_i sonically activated vibrational specific heat, μ_m the maximum relaxation absorption at $(f/p)_m$, $X = f/p$ and $X_m = (f/p)_m$.

Thus the first object of this work was to determine experimentally the value to be ascribed to the term $[(\eta)^{1/2} + (\gamma - 1)(K/C_p)^{1/2}]$ in eqn (2). This was done by measuring absorption in tubes of different radii. Also by varying f/p a test was made of the form of eqn (1) and experimental values for K_T and K_G were determined.

§ 2. Apparatus

The apparatus was a modification of one described previously (Lawley 1952). A small magnetostrictive rod was mounted in front of a narrow glass tube and a metal or glass reflector was driven in from the other end. The changes in anode current of the oscillator driving the rod were measured as the reflector moved along the tube. The natural logarithms of the differences between successive maximum and minimum values of anode current were plotted against the number of half wavelengths. Under certain conditions (see Hardy 1943) the intensity absorption μ per wavelength is given by the slope of the resulting graph.

It has been shown by Knudsen and Fricke that the frequency of maximum absorption f_m is very susceptible to slight changes in water vapour content (1% water vapour raises f_m from 20 kc/s to 2275 kc/s). Because of this it is difficult to obtain repeatable results in the region of molecular absorption when the gas sample is changed; hence in the work that follows no direct comparison was made between results obtained with different gas samples. The apparatus was therefore constructed so that three tubes, sources and reflectors could be mounted side by side and the whole interferometer system was enclosed in a gas-tight chamber which was kept at a constant temperature of 27°C. The gas was taken from a commercial cylinder, passed through calcium chloride drying tubes, and stored over phosphorus pentoxide for several days before being admitted to the interferometer chamber.

§ 3. Results

In order to interpret the experimental results for μ it is necessary to vary the relative values of the two terms in eqn (1) so that they can be evaluated separately. The first term may be found by varying r only and the second by varying X since different functions of X occur in the first and second terms.

For the method in which r was made variable three magnetostrictive sources having the same frequency were placed in three tubes of different radii mounted side by side. With the same gas sample filling the chamber the absorption was measured in the three tubes consecutively. K_T was then evaluated from the slope of the straight line graph of μ against $1/r$. An example of this is shown in figure 1. The straight line shows that the reciprocal radius law is correct and indicates an absorption due to viscosity and thermal conductivity 2% higher than that calculated from the theoretical equation. The intercept of figure 1 gives the value of the relaxation term in eqn (1), but this has little meaning since neither K_G nor X_m are known.

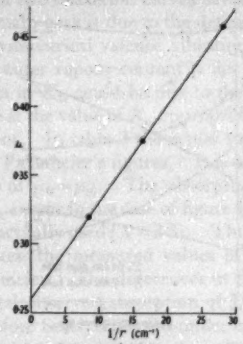


Figure 1. 40 kc/s, 2·34, 1·2, 0·7 mm diam tubes.

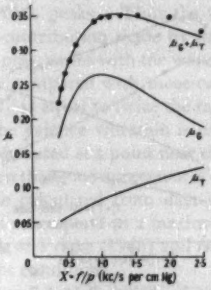


Figure 2. 40 kc/s, 2·34 mm diam tube.

Next, values of absorption coefficient were obtained for one of the tubes only as f/p was reduced through the range 130–20 cm Hg. In order to show whether the results are of the form of eqn (1), the values of the tube absorption μ_T only were first calculated from $\mu_T = K_T X^{1/2}$ using the value of K_T obtained from figure 1. Then the values of μ_T were subtracted from the experimental values of μ to give the absorption due to relaxation alone. A theoretical curve for the relaxation absorption $\mu_G = K_G X X_m/(X^2 + X_m^2)$ was then fitted to these values. This was done either by inspection where the peak of the relaxation curve was traced out or by plotting $X/(\mu - \mu_T)$ against X^2. The slope of this line gives $1/K_G X_m$ and the intercept X_m/K_G.

Figure 2 shows results obtained with one of the three tubes used previously. The experimental points are shown around the upper curve. The lower curve is a theoretical curve for tube absorption only and the middle one a theoretical relaxation curve μ_G. The upper curve is the sum of μ_T and μ_G, and this is to be compared with the experimental points. It appears from the graph that the form of eqn (1) is correct for both terms, but the work is not completely satisfactory because the contribution of each term was measured with different gas samples.

In the next experiment the two factors were determined at the same time. Two tubes of different radii and containing transducers of the same resonant frequency were placed side by side in the chamber. Absorption readings were taken in both tubes at each pressure throughout the range. Using the 1/radius

law (2) at each pressure a value of the viscosity–thermal conductivity constant was obtained, and from the mean of this K_T was calculated for each tube. As before, the values of $\mu_T = K_T X^{1/2}$ were subtracted from the experimental values of μ, leaving the absorption due to relaxation only. As was to be expected, the relaxation absorption was the same for both tubes since the same gas sample was used, and a theoretical relaxation curve was fitted to the points. Figure 3 shows the curves for μ_T for the two tubes, the common relaxation curve μ_G, and curves for the sum of μ_G and μ_T, which are to be compared with the experimental points in each case.

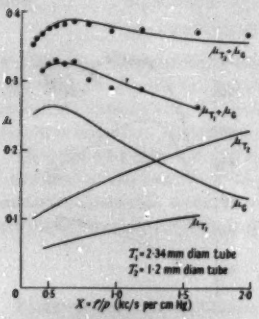


Figure 3. 40 kc/s, 2·34, 1·2 mm diam tubes.

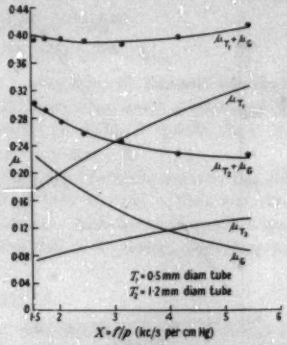


Figure 4. 120 kc/s, 0·5, 1·2 mm diam tubes.

In figure 4 an example is given of the results of a similar experiment with different tubes at 120 kc/s. In this case the values of f/p extend over a considerable range beyond the peak of the relaxation curve. In the results shown the slight effect of classical absorption in the free gas which was completely negligible at 40 kc/s has been previously subtracted from the experimental values of absorption. For the highest values of f/p, μ (classical) = 0·002.

The results are summarized in table 1.

Table 1

Fig. No.	Tube diam (mm)	Freq. (kc/s)	Range of X	X_m	K_G	K_T (exp) $-K_T$ (calc)*
1	2·34, 1·2, 0·7	40	0·534			+ 2
2	2·34	40	0·38–2·50	1·04	0·532	+ 2
3	2·34, 1·2	40	0·39–2·0	0·53	0·526	+ 4
4	1·2, 0·5	120	1·56–5·46	0·974	0·502	+13

* $[K_T \text{ (exp)} - K_T \text{ (calc)}]$ as % of K_T (calc).

§ 4. Discussion of Results

The results show that over a considerable range of f/p on both sides of the relaxation peak of CO_2 the viscous absorption of the tube and the relaxation absorption of the gas occur independently and are additive. No unusual effects

are observed in the term due to thermal conductivity and viscous drag at the walls, the observed absorption being a few per cent higher than that calculated theoretically. This is a similar result to that found in the case of classical gases.

The variation in X_m is considered to be due partly to slight differences in impurities in the gas samples, but mainly to the fact that the gases were not dried to exactly the same extent. The small variation in K_G is not at first to be expected since K_G does not vary as does X_m with small quantities of water vapour. This may be explained, however, by assuming the presence of a minor peak in addition to the main absorption peak. Pielemeier (1943) has obtained good agreement between the results of several previous workers and a theoretical curve which is the sum of two relaxation curves having values of K_G of 0·428 and 0·088 respectively. The main peak is due to the deformation vibrations while the minor one is due to the symmetrical valence vibration. Pielemeier shows that the variation of X_m with water vapour content is not the same for both peaks. Thus the variation shown in K_G could be due to the fact that the contribution made by the minor curve at the value of X_m appropriate to the main peak varies with the water vapour content. In table 2 the actual values of K_G are compared with those calculated from Pielemeier's figures. It is assumed that K_G is equal to twice the maximum value of $\mu_G = \mu_m$. The absorption μ' due to the valence vibration is calculated at X_m except in the case of figure 4, where it is calculated at a point near the values of X actually used ($X = 2{\cdot}5$). Thus it may be seen that good agreement is obtained between the measured values of K_G and those calculated from data given by Pielemeier. The differences in the last column correspond to a maximum error in total observed absorption of 3%. The work of Fricke (1940) and others has therefore been confirmed under entirely different conditions.

Table 2

Fig. No.	X_m	Main peak μ_m	μ'	K_G (calc) = $2(\mu_m+\mu')$	K_G (measured)	K_G (meas) $-K_G$ (calc)
2	1·04	0·214	0·040	0·508	0·532	0·024
3	0·53	0·214	0·044	0·516	0·526	0·010
4	0·974	0·214	0·021	0·470	0·502	0·032

Acknowledgments

The author wishes to thank Professor W. E. Curtis for providing the facilities for carrying out this research and Dr. E. G. Richardson who suggested the problem and gave valuable advice. He is also grateful to Professor J. F. W. Bell for many helpful suggestions.

References

Fricke, E. F., 1940, *J. Acoust. Soc. Amer.*, **12**, 245.
Hardy, H. C., 1943, *J. Acoust. Soc. Amer.*, **15**, 91.
Herzfield, K. F., and Rice, F. O., 1928, *Phys. Rev.*, **31**, 691.
Lawley, L. E., 1952, *Proc. Phys. Soc.* B, **65**, 181.
Pielemeier, W. H., 1943, *J. Acoust. Soc. Amer.*, **15**, 22.
Van Itterbeek, A., and Mariens, P., 1938, *Physica*, **5**, 153.
Zartman, I. F., 1949, *J. Acoust. Soc. Amer.*, **21**, 171.

The Velocity and Absorption of Sound in Aqueous Sugar Solutions

By A. W. PRYOR* and R. ROSCOE
Physics Department, King's College, Newcastle-upon-Tyne

Substance of paper read before the Society on 13th July 1953; MS. received 28th August 1953

Abstract. Measurements of the absorption of sound have been made by a pulse method due to Pinkerton, and measurements of the velocity by a pulse-interference method, in the frequency range 6 to 25 Mc/s. Solutions of sucrose, d-glucose and lactose have been investigated at various concentrations and temperatures. The velocity results confirm the conclusion, originally derived from viscosity measurements, that the sugar molecules have 'solvation envelopes' (attached layers of water molecules) which decrease in thickness as the temperature of the solution rises. The ratio of the actual absorption to the 'classical' absorption is found to be fairly close to the value of this ratio for water (3·1) at all concentrations, while elementary considerations suggest that it should decrease towards unity as the concentration increases. A theoretical treatment reveals, however, an additional source of energy dissipation at high concentrations which approximately accounts for the constancy of this ratio. This dissipation arises from the high local viscous stresses which result from the squeezing-out of water molecules from between the sugar molecules when the solution is compressed. Similar considerations must apply to pure associated liquids if these contain two different states of molecular arrangement, one crystalline and one disordered. On this account a value of about 2·25 for the ratio of actual to classical absorption is to be expected for such liquids, while the observed values for alcohols and water range from 1·64 to 3·10. Some doubt is thus cast upon the theory that the major part of the absorption in water is due to a structural relaxation process.

§1. Apparatus

The absorption measurements were taken by means of a pulse method in the frequency range 6–25 Mc/s. The method (Pinkerton 1949 a) is now so well known as to need no detailed description. The interferometer used is shown in figure 1. The liquid is contained in a brass tube, heavily chromium plated on the inside and surrounded by an electric heater thermostated to $\pm 1°$C. The aluminized quartz crystal is lightly greased with silicone grease and pressed, by means of a spring, against a collar in the bottom of the tube to form a liquid-tight seal. A brass plate backing the crystal is used as a live lead. As the interferometer is open to the air, changes in the water content of the solutions can occur, so each solution was used for only one set of measurements. It is estimated that an accuracy to about 5% was obtained in the absorption measurements.

The velocity measurements were taken with the same apparatus, using a pulse interference method. Short pulses of oscillations are gated from a continuously running oscillator and applied to the crystal. The echo signals received by the crystal are amplified, mixed with part of the original continuous

* Now at Underwater Research Laboratories, Department of Scientific and Industrial Research, Auckland, New Zealand.

oscillations, and detected. As the reflector is moved through each wavelength, the echo signal passes in and out of phase with the original continuous oscillations and the height of the echo pulse changes from positive to negative. More than a hundred wavelengths are counted in this way and measured on the accurate screw driving the reflector. The frequency is measured with a crystal-check, heterodyne wavemeter and the velocity determined to $\pm 0{\cdot}1\%$.

§2. Measurements

Measurements were made on aqueous solutions of sucrose, glucose and lactose. These were prepared respectively from crystalline 'Analar' cane sugar, crystalline anhydrous d-glucose of 'Reagent' purity, and crystalline lactose monohydrate (bacteriologically tested). The strengths of the solutions used are quoted as percentages by weight, that is to say the number of grammes of sugar dissolved in 100 g of solution.

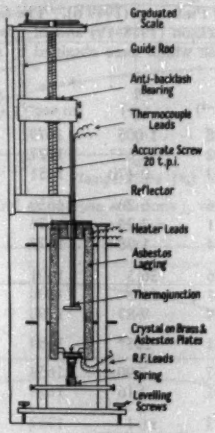


Figure 1. The interferometer.

Preliminary absorption measurements were made with 16·7, 28·6, 37·5 and 50% solutions of sucrose, at temperatures ranging from 15 to 80°C and at frequencies of 6, 12, 18 and 25 Mc/s. These measurements showed that the absorption coefficient α is directly proportional to the square of the frequency f over this frequency range. The measurements were then repeated more accurately at convenient frequencies (i.e. frequencies which gave absorptions of 1 to 5 dB cm^{-1}) and extended to include higher concentrations and other sugar solutions. Table 1 gives the final results for sucrose solutions at 20, 40 and 60°C, for glucose solutions at 20 and 40°C and for lactose solutions at room temperature.

Velocity measurements made on sucrose solutions of concentrations ranging from 16·7 to 80%, at temperatures from 11 to 81°C, are shown in figure 2. Those made on glucose solutions of concentrations ranging from 16 to 70%, at temperatures from 17 to 79°C are shown in figure 3, while those made on lactose solutions are shown in table 1.

§3. Density and Viscosity Data

Since the velocity of sound depends on the density of the solution as well as on its compressibility, an accurate knowledge of the density is necessary for the theoretical interpretation of the velocity observations. A comprehensive table of the density of sucrose solutions at 20°C is given in *International Critical Tables*. The values there shown agree, to within about 3 parts in 1000, with values calculated from the density of water and the dry sugar on the assumption that no volume change occurs as a result of solution. The densities of the glucose solutions used in this work were measured at 20°C with an accuracy to 1 part in 1000, and here again it was found that they agreed well with values calculated from the density of water and the density of the dry sugar. This method of calculation was therefore used to determine the densities of sucrose and glucose

Table 1. Absorption in Sugar Solutions

The experimental results are given in the form of the ratios α/f^2 and α/α_c. The values for water are those obtained by Pinkerton (1949 b). The figures for the viscosity are from the results of Bingham and Jackson (1918–19) and of Powell (1914), with the exception of those marked with an asterisk which were obtained by measurements using the falling sphere method.

	Temp. (°C)	ρ (g cm^{-3})	η (cP)	V (m sec^{-1})	$\alpha/f^2\times10^{15}$ (sec^2 cm^{-1})	$\alpha_c/f^2\times10^{15}$ (sec^2 cm^{-1})	α/α_c
Water 7·5–67·5 Mc/s	20	0·998	1·005	1479	0·253	0·082	3·10
	40	0·992	0·653	1527	0·146	0·0485	3·01
	60	0·983	0·470	1551	0·101	0·0336	3·02
40% sucrose 18 Mc/s	20	1·176	6·20	1650	0·97	0·309	3·1
	40	1·171	3·25	1674	0·49	0·156	3·1
	60	1·164	1·98	1677	0·27	0·095	2·8
60% sucrose 12 Mc/s	20	1·286	56·5	1801	5·88	1·97	3·0
	40	1·281	21·3	1800	2·10	0·750	2·8
	60	1·274	9·83	1790	0·93	0·356	2·6
69% sucrose 6 Mc/s	20	1·341	255*	1884	24·0	7·46	3·2
34·9% glucose 25 Mc/s	20	1·107	3·90	1654	0·77	0·205	3·7
	40	1·101	2·16	1678	0·35	0·109	3·2
49·3% glucose 12 & 25 Mc/s	20	1·223	11·1	1757	1·75	0·44	4·0
	40	1·219	5·09	1767	0·74	0·20	3·7
70% glucose 6 & 12 Mc/s	20	1·339	155*	1939	14·2	4·24	3·4
23·4% lactose 25 Mc/s	25	1·096	2·14*	1582	0·36	0·13	2·8
35% lactose 25 Mc/s.	21	1·143	4·30*	1633	0·71	0·23	3·1

solutions at higher temperatures. The densities of the lactose solutions, only required at room temperatures, were determined directly.

Considerable theoretical importance is attached to the ratio of the actual absorption coefficient α to the 'classical' absorption coefficient α_c calculated from Stokes' formula given in eqn (3) below. The latter involves the viscosity of the solution as well as its density. Extensive measurements of the viscosity of sucrose solutions, of concentrations up to 60%, have been made by Bingham and Jackson (1918–19) over a very wide range of temperatures, and Powell (1914) has made viscosity measurements on d-glucose solutions, of concentrations up to 49·33%, from 25 to 50°C. These cover all the solutions for which

absorption measurements were made, except 69% sucrose, 70% glucose, and 23·4 and 35% lactose.

The viscosity of these solutions was therefore determined by the falling sphere method. Steel ball bearings of 0·159 cm diameter were used in fairly narrow tubes (0·3 and 0·7 cm diameter) so that the velocity of fall was slow. In each case the apparatus was calibrated using a 40 or 60% sucrose solution.

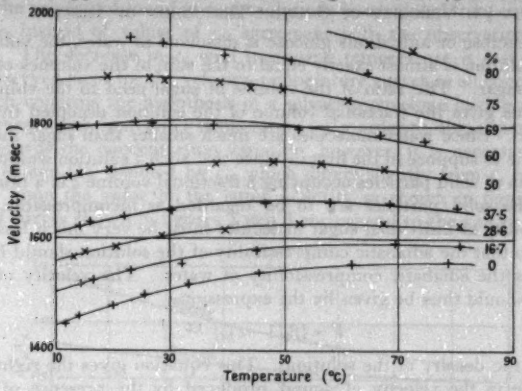


Figure 2. Velocity of sound in sucrose solutions : variation with temperature and concentration.

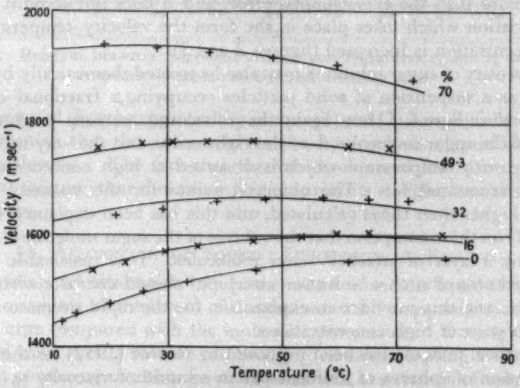


Figure 3. Velocity of sound in glucose solutions : variation with temperature and concentration.

Powell's figure for 49·33% glucose was checked by this method, and the values for the other solutions are included in table 1. As a check on the method, a further experiment was performed with 69% sucrose, using a wider tube (1·3 cm diameter), the absolute value of the viscosity being calculated directly from the Ladenburg formula (Hatschek 1928, p. 35). The result was within 3% of the value obtained by the comparative (narrow tube) method.

It is to be noted that, in the case of a 69% sucrose solution, an increase of 1% in the water content reduces the viscosity of the solution by 20%. It was therefore necessary to keep the sugars in stoppered bottles before preparing the solutions, and to avoid any change in water content during the preparation and use of the more concentrated solutions.

§4. Velocity of Sound: Discussion of Results

When sucrose or anhydrous glucose is dissolved in water, the volume of the resulting solution is almost exactly equal to the sum of the volumes of the water and solid sugar. The ratio of the volume of sugar used to the volume of the solution thus gives the fractional volume of the solution occupied by the sugar molecules c. Since water molecules are much smaller than sugar molecules, it is reasonable to suppose in the first instance that such a solution would behave as a suspension of solid particles occupying a fractional volume c in a homogeneous liquid. The solid particles are to be regarded as incompressible, since the compressibility of individual sugar molecules must be very much less than that of water, so that the adiabatic compressibility of the solution should be $\beta_0(1-c)$ where β_0 is the adiabatic compressibility of water. The velocity of sound in a solution should thus be given by the expression

$$V=\{\beta_0(1-c)\rho\}^{-1/2} \qquad \ldots\ldots(1)$$

where ρ is the density of the solution. This equation gives the right order for the increase in the velocity of sound produced by the presence of the sugar molecules. It gives numerical values, however, which differ from those observed by much more than the experimental error, and it does not account adequately for the variation which takes place in the form the velocity–temperature curves as the concentration is increased (figures 2 and 3).

The viscosity of sugar solutions may also be treated theoretically by regarding a solution as a suspension of solid particles occupying a fractional volume c in a homogeneous liquid. Here again the calculated increase produced by the presence of the sugar molecules is of the right order, but the very rapid decrease of viscosity with temperature which is observed at high concentrations is not adequately accounted for. The observed values for the viscosity are always somewhat higher than those calculated, and this has been explained by Einstein (1906, 1911) on the assumption that the volume of the sugar molecules is effectively increased by a layer of attached water molecules. It is reasonable to suppose that the thickness of such a 'solvation envelope' should decrease with increasing temperature, and this provides an explanation for the rapid decrease of viscosity with temperature at high concentrations.

A theoretical formula has been proposed by Roscoe (1952) for the viscosity η of a suspension of spheres of uniform size in a liquid of viscosity η_0:

$$\eta=\eta_0(1-1{\cdot}35c)^{-5/2}. \qquad \ldots\ldots(2)$$

This should be valid for medium and high concentrations, and in fact it agrees with experimental observations for spheres of very nearly uniform size. It should be applicable for ellipsoidal particles, provided their eccentricities are not very great, and may therefore be used for sugar solutions in order to calculate effective values of c from viscosity data. Such values of c, which will be denoted by the symbol c_2, represent the fractional volume of the solution occupied by the

solvated molecules, and are therefore larger than the values of c derived from the volume of dry sugar (which will be denoted by the symbol c_0). Values of the ratio c_2/c_0 for sucrose calculated from the viscosity data of Bingham and Jackson decrease from 1·23 to 1·06 as the temperature rises from 20° to 75°c. In the case of glucose, the data of Powell give values of this ratio decreasing from 1·16 to 1·03 as the temperature rises from 25° to 50°c.

When the velocity of sound in sugar solutions is calculated from eqn (1), using values of c_2 instead of values of c_0, agreement with the observations is much improved. The decrease in c_2 with increasing temperature is just such as to give the observed form for the velocity–temperature curves at all concentrations, a fact which confirms the hypothesis of a solvation envelope having a thickness which decreases with increasing temperature. In order to obtain perfect agreement as regards the concentration variation, however, it is necessary to use a value of c in eqn (1), denoted by c_1, which differs slightly from c_2 by amounts depending upon the value of the latter quantity as shown in figure 4. The continuous curves shown in figures 2 and 3 correspond to the values of c_1 given in figure 4.

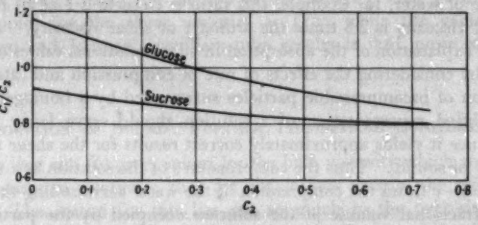


Figure 4. Relation between fractional volume concentrations c_1 and c_2 calculated respectively from velocity and viscosity data.

It is to be noted that the applicability of eqns (1) and (2) is essentially dependent on the assumption that the water behaves as a homogeneous liquid surrounding the sugar molecules. The fact that the values obtained for c_1 and c_2 are not exactly equal suggests that this assumption represents a slight over-simplification of the problem.

§5. Absorption of Sound: Discussion of Results

In discussing the absorption of a plane sound wave in a liquid it is necessary to consider a volume of the liquid of dimensions small compared with the wavelength but large compared with the molecular dimensions. During the passage of the sound wave such a volume can be regarded as undergoing a homogeneous deformation consisting of a change of shape as well as a change of volume. The viscous stresses which come into operation as a result of the change in shape give rise to a dissipation of energy and hence an absorption of the sound. The resulting absorption coefficient is given by the 'classical' formula due to Stokes:

$$\alpha_c = \frac{8}{3}\pi^2\frac{\eta f^2}{\rho V^3} \qquad \ldots\ldots(3)$$

where η is the viscosity of the liquid, ρ its density, V the velocity of sound and f its frequency.

In general there is also a dissipation of energy associated with the compressional part of the deformation, which results from a lag of the compression behind the pressure. At low frequencies the effect of such a lag is to increase the mean pressure above the normal value s/β (where s is the compression and β the adiabatic compressibility) by an amount proportional to the rate of compression. That is to say it is increased by an amount $\mu\, ds/dt$, where μ is a constant known as the bulk viscosity of the liquid. The extra dissipation of energy, resulting from the presence of this additional stress, raises the absorption coefficient from the classical value given by (3) to

$$\alpha = \frac{8}{3}\pi^2 \frac{\eta + \frac{3}{4}\mu}{\rho} \frac{f^2}{V^3} \qquad \ldots\ldots(4)$$

so that the ratio of the actual absorption to the classical absorption is given by

$$\frac{\alpha}{\alpha_c} = 1 + \frac{3}{4}\frac{\mu}{\eta}. \qquad \ldots\ldots(5)$$

In the case of water, for example, this ratio is close to 3·1 at all temperatures, so the bulk viscosity is 2·8 times the ordinary or shear viscosity.

For the explanation of the absorption in sugar solutions, values of μ/η may be estimated by considering the effects of rate of compression and rate of shear on a suspension of incompressible particles surrounded by a homogeneous liquid. This simplified representation of a solution should serve fairly well for the purpose, since it yields approximately correct results for the shear viscosity and the velocity of sound. Thus the compression s of the solution as a whole should be equal to $1-c$ times the compression of the water surrounding the particles, c being the fractional volume of the solution occupied by the particles. Then for a given rate of compression ds/dt the pressure should exceed the normal value by an amount equal to the product of the bulk viscosity of water μ_0 and the rate of compression of the water $(ds/dt)/(1-c)$. Thus the solution should have a bulk viscosity

$$\mu_1 = \mu_0/(1-c). \qquad \ldots\ldots(6)$$

This formula gives a bulk viscosity which increases with concentration much less rapidly than the shear viscosity which is given by eqn (2). Consequently, when it is used with eqn (5) it yields values of α/α_c which decrease from the value 3·1 (for pure water) towards unity as the concentration increases, while the experimental values remain fairly close to 3·1 for all concentrations. The theory therefore fails badly at the higher concentrations, as may be seen by comparing the third and fourth columns of table 2.

Table 2. Observed and Theoretical Values of α/α_c for Concentrated Sugar Solutions

Concentration (% by mass)	Temp (°C)	Obs	$1+\frac{3}{4}\frac{\mu_1}{\eta}$	$1+\frac{3}{4}\frac{\mu_1+\mu_2}{\eta}$
Sucrose 60	20	3·0	1·07	2·32
Sucrose 60	40	2·8	1·12	2·37
Sucrose 60	60	2·6	1·17	2·42
Sucrose 69	20	3·2	1·02	2·27
Glucose 70	20	3·4	1·03	2·28

In this table the values of μ_1 have been calculated from eqn (6) using values of c corresponding to the velocity data c_1. The values of η used in calculating both the observed and theoretical values of α/α_c are those obtained from capillary-tube and falling-sphere methods. Some justification is required for the use of shear viscosity coefficients determined by such methods, since a concentrated solution might well exhibit different values of η when flowing continuously and when subjected to small oscillatory deformations. Thus, when a concentrated suspension of solid particles is continuously sheared, a considerable temporary aggregation of the particles must occur, but there is no reason to suppose that this will happen as a result of the very small strains occurring in the passage of a sound wave; and the viscosity may well depend upon the spatial distribution of the particles. It is possible, however, to obtain values of the shear viscosity for small oscillatory deformations from measurements on the absorption of shear waves; and observations made with a 60% sucrose solution show that there is little difference, if any, between this quantity and the shear viscosity for steady flow. These observations, which were made with a torsionally oscillating transducer at 50 kc/s using the method of McSkimin (1952), gave a value of $0{\cdot}64 \pm 0{\cdot}06$ poise for the shear viscosity at 17°c, while the value obtained from Bingham and Jackson's capillary-tube results is 0·61 poise at the same temperature.

§6. Absorption of Sound: Further Theoretical Considerations

It is clear that eqn (6) gives values for the bulk viscosity which are much too low for concentrated solutions: they are not, in fact, even of the right order of magnitude. The reason for this lies not so much in the over-simplification involved in treating the solution as a suspension of incompressible solid particles in a homogeneous liquid as in a serious omission which has been made in applying that treatment. This may be made clear in general terms as follows: When such a suspension is compressed, the particles are brought closer so that the surrounding liquid suffers local changes in shape as well as compression. In particular, as two close particles approach, the liquid between their points of near-contact is squeezed out, and this gives rise to high viscous stresses which contribute to the compressional resistance of the suspension, a factor which has been neglected in the derivation of eqn (6).

In the experiments, the frequencies used were very much less than the reciprocal of $\beta_0\eta_0$ (the product of the compressibility and viscosity of the liquid surrounding the particles). When this condition holds, the deformation of this liquid may be separated into a homogeneous compression $s/(1-c)$ without change of shape and a flow with change of shape but without compression obeying Stokes' equations for the flow of an incompressible fluid. The total rate of dissipation of energy in unit volume of the suspension $\mu(ds/dt)^2$ is the sum of the rates of dissipation which would occur if these motions took place separately. It may therefore be written as

$$\mu\left(\frac{ds}{dt}\right)^2 = \mu_1\left(\frac{ds}{dt}\right)^2 + \mu_2\left(\frac{ds}{dt}\right)^2 \qquad \ldots\ldots(7)$$

where μ_1 is given by eqn (6), and μ_2 is a quantity dependent upon η_0 (but not upon μ_0), which is associated with the incompressional part of the motion. It

may be regarded as the contribution of the latter to the total bulk viscosity μ of the suspension. Thus eqn (5) may be rewritten as

$$\frac{\alpha_c}{\alpha} = 1 + \frac{3}{4}\frac{\mu_1 + \mu_2}{\eta}. \qquad \ldots\ldots(8)$$

A calculation of the ratio μ_2/η is therefore necessary for a proper theoretical deduction of the value of α/α_c.

At very low concentrations μ_2/η can be shown to be equal to $4c/3$, but in such cases it is of little importance as it is negligible in comparison with μ_1/η, which approaches the value μ_0/η_0 (2·8 for water) as the concentration approaches zero. At intermediate concentrations a calculation of μ_2/η would be very difficult. At high concentrations, however, it can be obtained quite simply by making a comparison between (*a*) the energy dissipation resulting from the incompressional part of the motion of the liquid when the suspension as a whole is subjected to compression, and (*b*) the energy dissipation which takes place when the suspension is subjected to pure shear.

When a concentrated suspension is compressed, the liquid suffers its greatest change of shape where it is squeezed out from between the points of near-contact of adjacent particles. Now the rate of dissipation of energy taking place in a liquid which is squeezed out from between two circular flat plates of radius a and instantaneous separation h, moving together with velocity v, is $3\pi\eta_0 a^4 v^2/2h^3$, an expression which increases without limit as the separation h is decreased. Thus in a concentrated suspension where adjacent particles are nearly touching, all but an inappreciable fraction of the dissipation of energy takes place in the regions close to the points of near-contact, as indicated in figure 5 (*a*).

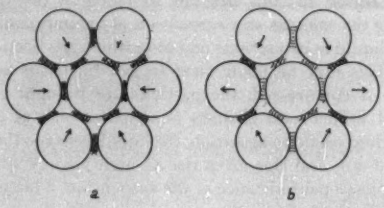


Figure 5. Showing the regions of high rate of shear in a suspension of solid particles undergoing (*a*) compression and (*b*) pure shear.

Similar considerations apply when such a suspension is subjected to a pure shearing motion, consisting of a small increase of its dimensions in one direction accompanied by an equal decrease of its dimensions in another direction at right angles. It can be seen from figure 5 (*b*) that there is again a squeezing out of liquid between certain points of near-contact, but an inflow of liquid between some others. A complication is introduced here by the fact that there is generally an appreciable tangential component of the relative motion between adjacent particles. However, the rate of dissipation of energy between two circular flat plates of radius a which move parallel to one another with relative velocity v at a constant separation h is $\pi\eta_0(a^2v^2/h)$, and the ratio of this to the rate of dissipation for an equal normal motion of the plates is $2h^2/3a^2$, which tends to zero

if the separation of the plates is reduced indefinitely. Consequently the rate of dissipation due to the tangential component of the relative motion between adjacent particles is negligible compared with that due to the normal component of this motion in the case of very concentrated suspensions.

Thus if a concentrated suspension is subjected successively to compression and to pure shear, the ratio of the rates of dissipation of energy between a particular pair of particles is simply as the square of the ratio of the rates of approach of their surfaces in the two cases. For the purpose of calculating this ratio it may be sufficiently accurate to take the rate of approach of the surfaces (in either deformation) as being equal to the rate of approach which the centres of the particles would have if they were points in a continuum subjected to the same macroscopic deformation as the suspension. (This is certainly true for the case of a suspension of spheres, such as that shown in figure 5, in which each is symmetrically surrounded by neighbours.)

During a compressional deformation, the dimensions of a cube of the suspension, originally of unit side, contract at a rate r equal to $\frac{1}{3}ds/dt$. For comparison (figure 6), a case of pure shear is considered in which one side of the same cube is contracting at the rate r, another is extending at the same rate, while the third side (at right angles to the plane of the paper in figure 6) remains

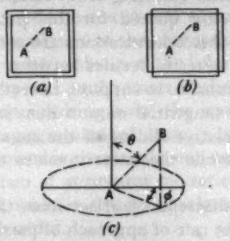


Figure 6. Continuum undergoing (*a*) compression and (*b*) pure shear, and (*c*) system of angular coordinates.

unchanged. Now if A and B are any two points fixed in a continuum undergoing these two deformations successively, the square of the ratio of their rates of approach in the two cases is as $1:(\cos^2\theta-\sin^2\theta\cos^2\phi)^2$, where θ and ϕ are angular coordinates taken as in figure 6.

Now the adjacent pairs of particles in the suspension can be divided into groups according to the relative positions of the two particles forming the pair. In each such group there will be a large number of pairs, and the orientation of the line joining their centres will be randomly distributed. In compression, the rates of dissipation of energy between all pairs in a group will be the same, but in simple shear they will differ on account of the dependence of the rate of approach on the orientation of the line of centres. Thus for the whole group the ratio of the rates of dissipation of energy in pure shear and in compression is obtained by averaging $(\cos^2\theta-\sin^2\theta\cos^2\phi)^2$ over all orientations of the line of centres. The resulting value 4/15 is of course the same for all groups, and is therefore the ratio of the rates if dissipation in pure shear and compression for the whole suspension.

This ratio is also obtainable in terms of the macroscopic constants of the suspension μ_2 and η: the part of the dissipation of energy in compression which has been considered here is given by eqn (7) as $\mu_2(ds/dt)^2$, or $9\mu_2 r^2$, for unit volume. On the other hand, the rate of dissipation of energy per unit volume in the case of pure shear is given by ordinary hydrodynamics as $4\eta r^2$, so that $4\eta r^2/9\mu_2 r^2 = 4/15$, and hence for a concentrated suspension

$$\mu_2/\eta = 5/3. \qquad \ldots\ldots(9)$$

This value of μ_2/η has been used with eqn (8) to calculate the theoretical values of the ratio α/α_c for concentrated sugar solutions shown in the last column of table 2. The agreement between calculation and experiment, though not good, is now much improved. It is evident that the theory gives a contribution μ_2 to the bulk viscosity μ which is of the right order of magnitude, although always somewhat low.

The cause of the residual discrepancy does not appear to lie in the assumption that sugar solution can be treated as a suspension of particles surrounded by a homogeneous liquid. Thus, in the derivation of eqn (9), homogeneity of the surrounding liquid has only been assumed in order to show that the energy dissipation resulting from relative tangential motion of particle surfaces is negligible compared with that resulting from relative normal motion, classical hydrodynamic formulae being quoted for this purpose. Now in the sugar solutions quoted in table 2 the concentrations are so high that there can seldom be more than one layer of water molecules between the near-points of adjacent sugar molecules. It is reasonable to suppose, nevertheless, that even here much less energy is dissipated in tangential motion than in normal motion, since the former only involves the relative sliding of the sugar molecules separated by a layer of water molecules, while the latter involves the squeezing out of water molecules.

It is probable that the discrepancy arises from the assumption made in the derivation of eqn (9) that the rate of approach of particle surfaces can be derived from the rate of approach of points fixed in a continuum undergoing the same deformation as the suspension. This may only be very approximately true when the particles depart appreciably from the spherical form, and it is to be noted that the length of the unsolvated sucrose molecule is nearly three times its breadth, while the glucose molecule is somewhat closer to the spherical form.

§7. The Absorption of Sound in Associated Liquids

The theory developed above for the explanation of the absorption of sound in concentrated sugar solutions has some bearing upon the absorption in pure associated liquids. In classifying liquids according to their absorption characteristics Pinkerton (1949 b) has noted that the associated liquids n-amyl alcohol, n-butyl alcohol, n-propyl alcohol, ethyl alcohol, methyl alcohol and water have values of the ratio α/α_c falling within the limited range 1·64–3·10. Other liquids have values of this ratio either very close to unity or else many times greater. The high absorptions observed in the latter cases can be explained on the basis of a thermal relaxation mechanism.

In the case of water, however, the excess of the absorption over the 'classical' value cannot be explained by a thermal relaxation mechanism (Fox and Rock 1946). Structural relaxation, resulting from molecular rearrangement under

pressure, has been proposed as an explanation in this case, and a detailed theory has been worked out by Hall (1948). It is supposed that two different states of molecular arrangement exist in ordinary water: in state 1 the structure is ice-like and the compressibility is low, while state 2 is disordered and has high compressibility, but the density of state 2 is greater than that of state 1. This compression should cause a relative increase in the number of molecules in state 2, and if this change is not instantaneous the compression will lag behind the pressure change so that the liquid will exhibit a bulk viscosity. By making a number of additional assumptions, Hall has calculated values of this bulk viscosity which, used with eqn (4), give results for the absorption coefficient in good agreement with the observations over the whole temperature range.

Now since water flows continuously under shear stress, the molecules in the ice-like state 1 must be regarded as forming separate aggregates surrounded by molecules in the mobile state 2. It is to be noted that in Hall's theory the greater part of the molecules are in state 1, and if this is the case water must itself be regarded as a highly concentrated suspension of relatively incompressible solid particles in a liquid. On this account alone, water is to be expected to exhibit a bulk viscosity roughly equal to 5/3 of its shear viscosity, in accordance with the theory developed in the previous section. From eqn (5) this is seen to correspond to a ratio α/α_c equal to 2·25 as compared with the observed ratio of about 3·1. It is therefore to be inferred that if structural relaxation does occur it is only responsible for a small part of the absorption.

The major part appears to be made up of the 'classical' absorption due to the macroscopic shear strains imposed on the water by the passage of the sound wave, together with the absorption resulting from the squeezing-out of molecules in state 2 from between the aggregates of state 1 molecules on compression. Although the latter mechanism is essentially one of molecular rearrangement, it differs from the mechanism of structural relaxation in that the changes only involve molecules in state 2, while in structural relaxation the molecules undergo transitions between states 1 and 2.

The results for the other associated liquids can be accounted for quite well by the same assumption as to the existence of states 1 and 2 (with a preponderance of state 1), again without introducing the hypothesis of structural relaxation. Here the observed values of α/α_c range from 1·64 to 2·64 as compared with the theoretical value 2·25.

ACKNOWLEDGMENTS

One of the authors (A.W.P.) wishes to thank the Australian Department of Supply for the provision of a Scholarship, and Dr. E. G. Richardson for his interest and advice on the experimental part of the work.

REFERENCES

BINGHAM, E. C., and JACKSON, R. F., 1918–19, *Bull. U.S. Bur. Stand.*, **14**, 59.
EINSTEIN, A., 1906, *Ann. Phys., Lpz.*, **19**, 289 ; 1911, *Ibid.*, **34**, 591.
FOX, F. E., and ROCK, G. D., 1946, *Phys. Rev.*, **70**, 68.
HALL, L., 1948, *Phys. Rev.*, **73**, 775.
HATSCHEK, E., 1928, *The Viscosity of Liquids* (London : Bell).
MCSKIMIN, H. J., 1952, *J. Acoust. Soc. Amer.*, **24**, 355.
PINKERTON, J. M. M., 1949 a, *Proc. Phys. Soc.* B, **62**, 286; 1949 b, *Ibid.*, **62**, 129.
POWELL, C. W. R., 1914, *J. Chem. Soc.*, **105**, 1.
ROSCOE, R., 1952, *Brit. J. Appl. Phys.*, **3**, 267.

REVIEWS OF BOOKS

Exercises in Experimental Physics, edited by N. C. B. ALLEN and L. H. MARTIN. Pp. xiii + 237. (Melbourne: University Press, 1952). 30*s*.

This work is a collection of experiments designed to supplement the second-year courses of physics in the University of Melbourne. The experiments described would be suitable in general for an ordinary B.Sc. course or for the less advanced parts of an honours course in Universities in England.

The book is described as being essentially the sixth edition of *Practical Physics*, in which the laboratory exercises were assembled by Professor Laby and the text was prepared by members of the staff of the Physics Department of Melbourne University.

The fact that the book has been written especially to conform to the needs of a particular group of students of necessity limits its usefulness as a textbook for other courses which may extend beyond its bounds. But the same remark would apply to any such work of reasonable size.

A more serious limitation is that it is written for undergraduates at a particular stage in their studies. This gives the appearance of omission and lack of balance. Thus, in a brief account of experiments on elasticity, nearly one half of the space is devoted to the interference method of measuring Poisson's ratio. The chapters on the measurements of inductances and capacitances are brief and incomplete and the necessity for shielding, the use of non-inductive resistances and pure harmonic oscillators could with advantage be more strongly emphasized.

On the other hand there are many useful hints scattered throughout the book, such as those on the choice of a galvanometer.

An important feature is the inclusion of a chapter on observations and statistical theory, with two experiments as illustrations.

The properties of vacuum tubes, the construction of a power supply unit and the use of the Geiger counter are described and illustrated in a well chosen series of experiments, designed to explain the principles of these instruments without emphasis at the expense of other fundamental parts of this course.

H. T. FLINT.

Progress in Metal Physics, Volume 3, by B. CHALMERS (Ed.). Pp. viii + 334. (London: Pergamon Press, 1952.) 48*s*.

The regular appearance of these volumes, each containing up-to-date accounts of many aspects of metal physics, shows how successfully and rapidly the subject is now developing. The volume with which this notice is concerned, number three of the series, contains articles on a wide variety of topics. These may be roughly classified as, first, those dealing with the equilibrium properties of more or less perfect metallic crystals, i.e. electrical, structural and magnetic properties, and, secondly, those dealing with the properties of the secondary structure (crystal boundaries etc.) and crystallographic transformations. The articles in the first category are written against a background of a highly developed theory: the electron theory of metals based on wave mechanics. In the second category

the articles are largely descriptive and empirical, and where theory is introduced it is phenomenological rather than fundamental. To any reader who attempts to study the whole book this difference between the two main branches of the subject quickly makes itself felt. Chapters 2, 3 and 5, which deal respectively with the properties of metals at low temperatures, recent advances in the electron theory of metals, and ferromagnetism, have a certain coherence which enables the reader to see relations between the observational data discussed and to view each subject as a whole. The articles dealing with what the editor refers to as the "dynamic effects", on the other hand, call for a disciplined effort on the reader's part to assimilate the large number of individual observations, the relations between which are not always easy to see. In the absence of a firm theoretical background the selection of the definitive observations is difficult to make and the authors are compelled to record faithfully all the experimental data available, and this may result in an article such as, for example, that dealing with x-ray diffraction in strained metals, which reads almost like the minutes of some committee meeting.

In a short notice of this kind it is not possible to deal with each of the eight articles individually, but special mention may be made of one or two. The book opens with an excellent account of the crystallography of transformations by J. S. Bowles and C. S. Barrett. This article deals with martensite transformations and the atomic movements which take place in these phase changes. Although the account is clear and complete, the reader new to the subject will find that a considerable effort is called for to follow the details of the geometry of the different transformations which have been proposed to explain the observational data. The physical (as distinct from the purely crystallographic) nature of the process is also discussed, but it cannot be said that the theory, as presented in this article, is very satisfactory; many gaps, according to the authors, remain to be filled.

Two other chapters which may be specially mentioned are Chapters 3 and 7. Chapter 3 deals with recent advances in the electron theory of metals and is written by N. F. Mott. As indicated by the title, this is not a critical account of established theory, but largely a description of newly developed ideas. It is, perhaps, too much to hope that all of these, or even the greater part of them, will survive without substantial modification in the future. Chapter 7, by J. E. Burke and D. Turnbull, is concerned with recrystallization and grain growth. A clear account of experimental results, expressed in terms of the concepts of nucleation and growth, is followed by some pages of physical theory dealing with the kinetics of boundary migrations. This admirable article of some 70 pages can be comfortably read without prior knowledge of this particular subject.

One minor criticism of the production of the whole book is that the proof reading does not appear to have been done with sufficient care. For example, in the article just referred to the letters t and τ are used rather indiscriminately although they are intended to have quite distinct meanings. More remarkable, perhaps, is the slip in the editor's foreword where we find the word 'phrases' where 'phases' is intended. These, however, are small blemishes which do not seriously detract from the great value of this volume. Some idea of the extent of the information contained within its covers is indicated by the fact that there are altogether 770 references to published papers, all of which are dealt with in the text.

H. JONES.

Calculus of Variations with Applications to Physics and Engineering, by ROBERT WEINSTOCK. Pp. x+326. (London: McGraw-Hill, 1952.) 55*s*. 6*d*.

Many of the laws of mathematical physics can be set up in two alternative ways; a law may be formulated either as a partial differential equation or as the necessity for a definite integral to have its extreme value. The first approach is the more common one, so much so that it has often been said that the aim of physics is to express all its laws in this form. The second method is less easy to understand but has the advantage of including the boundary conditions directly. Also, where a solution in terms of known functions is impossible, it leads to approximate methods which are often useful and are particularly appropriate for use with high-speed computing machines.

This book deals with both the mathematical theory of the calculus of variations and its application to a selection of problems in physics. These include examples on the dynamics of particles, electrostatics and geometric optics, but the main emphasis is on eigenvalue problems for which the associated differential equation is of the Sturm–Liouville type. The general treatment is clear and the amount of prior mathematical knowledge required by the reader is reasonably limited. A set of examples at the end of each chapter, with hints on their solution, fill in the details of proofs which are curtailed in the text and apply the results to further problems.

The physics student might be slightly misled at a few places. Thus the author gives the impression that Schrödinger originally formulated his equation independently of de Broglie's work. Also in the application of the Ritz method to the helium atom no physical explanation is given for the device of treating the nuclear charge as an adjustable parameter. These minor criticisms will not reduce the value of the book, especially to students of theoretical physics.

J. C. B.

Advanced Mathematics in Physics and Engineering, by A. BRONWELL. Pp. xvi+475. (London: McGraw-Hill, 1953.) 51*s*.

There is at present no lack of books on mathematics written specifically for engineers and physicists. Professor Bronwell, writing in language familiar to the engineer, gains much by avoiding the more ambitious and formal nature of many of these, and his book is, on balance, a worthwhile addition. The author does claim "a moderate balance of mathematical rigour", but this seems to rest mainly on a liberal use of such terms as "continuous", "differentiable", "regular", and "assuming it exists". Since these are not particularly well defined in the text, if at all, it is unlikely that the average reader will take much notice of them.

Choice of material is interesting and varied and the presentation is very clear apart from a certain looseness of language (e.g. "we can put $\partial/\partial z = 0$"). Layout and typography are excellent and only a few minor misprints were noted. Many problems are worked out in the text and a set of examples, with answers, is provided after each chapter. The main topics discussed are infinite series, Fourier series and integrals, solution of differential equations including Bessel's equation and Legendre's equation, vector analysis, functions of a complex variable, Laplace transforms, vibrations, hydrodynamics, heat flow, electromagnetic theory and roots of polynomials and stability criteria.

The work on convergence of series leaves much to be desired. Little is given in the way of proofs and, in the course of the proof that absolute convergence implies convergence, the author implicitly assumes what he is supposed to be proving. Other weaknesses are the statement of Taylor's theorem given, and the suggestion that convergence of series of complex terms should be examined by separation into real and imaginary parts. In fact nowhere in the book is the convergence of complex series treated on its own merits and nothing is said about even the simplest of tests for convergence of integrals although the need for such convergence is frequently mentioned later on. The various cases that arise in the series solution of differential equations are very clearly explained and the account of Bessel functions is particularly good. Legendre functions, however, receive much less attention and only one recurrence relation is given.

The chapter on vector analysis is, in general, quite adequate. The cartesian components of the curl of a vector are defined by means of a certain line integral, but the question as to whether the quantity so defined is in fact a vector (i.e. that the components retain their form on change of axes) is not raised. The section on Gauss' theorem is not very convincing. In the next chapter, equation (4) is certainly not Schrödinger's equation, as stated.

The chapters on hydrodynamics, heat flow and electromagnetic theory constitute attractive introductions to these subjects, but time differentiation following the motion, which students usually find difficult, is dismissed too lightly and there is a mistake on p. 283 where it is stated that, in steady irrotational flow, the constant in Bernoulli's equation varies from one streamline to another. The section on functions of a complex variable is clearly presented. It consists mainly of the theory of residues, contour integration and simple conformal transformations. This chapter ends with the Schwarz–Christoffel transformation, proved for a closed polygon. In the example given on a semi-infinite strip the reader is left to wonder what happens to the vertex at infinity.

The final chapter contains an elementary account of Laplace transforms and the Mellin inversion integral. An example of the latter involving a branch point would not have been out of place. A. N. GORDON.

Physiology of Vision (Vol. 9, no. 1, 1953, of *British Medical Bulletin*), by W. S. Stiles (Scientific Editor). Pp. ix + 80. (London: Medical Dept., British Council, 1953.) 15*s*.

This number of the *British Medical Bulletin* contains papers by British workers engaged in research on the physiology of vision, each author giving a general exposition of his field of study. The contents are as follows:

Introduction, by Sir John Parsons; Some problems concerning the formation and circulation of the aqueous humour, by Hugh Davson; Factors controlling ocular accommodation, by E. F. Fincham; The nature of refractive errors, by Arnold Sorsby; The interpretation of spectral sensitivity curves, by H. J. A. Dartnall; Chromatic aberration and the polychromatic theory, by Hamilton Hartridge; Defective colour vision, by W. D. Wright; Visual properties studied by subjective measurements on the colour-adapted eye, by W. S. Stiles; Stimulation of the retina with light fields of small size, by L. C. Thomson; Colour

vision in the peripheral retina, by R. A. Weale; The absolute sensitivity of the eye and the variation of visual acuity with intensity, by M. H. Pirenne; Electric records from the vertebrate optic nerve, by W. A. H. Rushton; Physiology of vision: a bibliography of British books and periodicals, by Katharine Tansley.

There are too many papers for all to be discussed adequately here, while it would be invidious to select some of them for review. They provide a great deal of information on the present state of the physiology of vision. This field is now engaging the attention of an increasing number of scientists all over the world and seems to be entering upon a new period of progress and development similar to that of the time of Helmholtz. The non-specialist may find that some of the contributions do not make very easy reading, but this is only to be expected from the nature of the subject matter which involves intriguing problems in modern physiology, physics and chemistry. The remark that "it is easier to say something new than to reconcile things already said" is particularly applicable to the study of visual problems. In this connection, the paper by the Scientific Editor, Dr. W. S. Stiles, should be mentioned for its critical study of the logical relationships between the different lines of approach in the study of human colour vision, including the striking contributions made by the author himself to this difficult field.

M. H. PIRENNE.

The General Properties of Matter, by C. J. SMITH. Pp. viii + 580. (London: Edward Arnold, 1953.) 50*s*.

There is probably no such thing as an objectively just price. A king was willing to offer a kingdom for a horse. No doubt, too, there will be students who will have so perceptive an appreciation of the value of book-learning that they will be overwhelmed at their good fortune in getting all that Dr. C. J. Smith and Messrs. Edward Arnold & Co. offer them for a mere fifty shillings, and even congratulate themselves that they will need several other volumes before they are fully equipped for their studies. The majority, however, will make unfavourable comparisons between the high finance of the publisher and the economy of those from whom they draw their grant. Without laying claim to detailed knowledge of publishing costs (which are indeed high), it surely can be said that it ought to be possible to produce the same service to the student at half the price. If the publishers really cannot do it, then there is surely a case here for a radical change of system. Books bought at so high a price tend to be sold by the student as soon as he is finished with them for examination purposes. This is doubly to be deplored: in the first place such books should form the foundations of the student's future library; in the second place it cuts down sales and thereby keeps the price up.

Let it now be said that, working within the accepted frame of reference, both author and publisher are in every way to be congratulated. This book can be confidently recommended. The syllabus for the B.Sc. General Degree is well covered in a lucid style, supported by instructive diagrams, which gives us an appetite for the forthcoming volumes, dealing with electricity (now in the press), optics, atomic physics and electronics.

The most popular combination of subjects for the General Degree is Physics, Pure Mathematics and Applied Mathematics, which overlap and supplement each other admirably to give a unified set of studies. The physics subjects,

however, make early demands on the student's mathematics which force the physics teacher to anticipate what will later be covered in greater detail in mathematics lectures, and, of course, there are many students who rely entirely on their physics tutors for their mathematics. With these problems in mind, Dr. Smith very sensibly starts immediately with an introduction to vector notation, Fourier analysis, differential equations and one or two other essential mathematical topics, which are succinctly dealt with. Separate chapters cover particle dynamics, rigid dynamics and simple harmonic motion in sufficient detail for those who are not reading Applied Mathematics as a separate subject.

The chapters on intensity of gravity and gravitation are commendable. Gravitational potential is defined as " the work per unit mass done against the field in bringing up a small mass from infinity to the point considered ", a definition consonant with the customary approach in electrostatics and magnetism. As in the rest of the book, Dr. Smith constantly gives the student a sense of historical perspective.

There are good chapters on elasticity and the bending of beams and springs, which are notable for their generosity of practical details, both with regard to experimental procedures and concerning the less formalized aspects of the physical behaviour of materials subject to stress.

The compressibility of liquids, solids and gases might well serve as a model for textbook writers, as dealt with by Dr. Smith. His material is everyone's material, but Dr. Smith makes it read almost as a story which carries the reader along. It is perhaps questionable how many students will actually take up the many references in the footnotes to original papers : e.g. Oersted, 1827, *Pogg. Ann.*, **9**, 603, or *Mém. de l'Acad. Fran.*, 1847, **21**, 330. But it is as sure as death and taxes that this is the right way to approach students at this stage, giving them a taste of and a taste for reference to sources of justification rather than *ex cathedra ipse dixit*. The style of the whole book is adult and the English exemplary, particularly for the many present-day students who have been taught in schools where carelessness of self-expression is no disgrace.

The remaining chapters of the book deal with surface tension, fluid and solid flow, osmosis, diffusion and vacuum practice, maintaining a high standard throughout. There is an ample selection of examples following each chapter. The author leaves many of these without a stated answer, believing that at this stage the student should acquire an attitude of mind which says ' If this solution isn't correct, none is '. Where answers are given, he says he almost dares hope that some of them are wrong, to frustrate the student who proceeds by blind substitution in a formula. There is a comprehensive index which leaves the student with little excuse for not being able to find what he needs very quickly. The format is a credit to printers and publishers.

The price is a great pity; it is to be hoped it does not deter any tutor from recommending this book or any student from buying it, for it is one of those books that ought not to be sold as soon as the examination is passed.

M. J. MORONEY.

CONTENTS OF SECTION A

PAGE

Editorial 1

Mr. B. SCHIFF. A Calculation of the Eigenvalues of Electronic States in Metallic Lithium by the Cellular Method 2

Dr. N. H. MARCH. Momentum Distribution of Electrons in Solids: Results for some Metals using the Thomas–Fermi Method 9

Mr. L. S. KOTHARI. Riesz Potential and the Elimination of Divergences from Quantum Electrodynamics 17

Dr. B. L. MOISEIWITSCH. On the Electron Affinities of Atomic Fluorine, Oxygen and Lithium 25

Dr. M. R. SCHAFROTH. Self-Consistent Spin-Wave Theory for the Ferromagnetic Exchange Problem 33

Dr. L. A. RADICATI. Isotopic Spin and Coulomb Forces: II—Excited States of Light Nuclei 39

Mr. N. K. BRIDGE and Dr. H. G. HOWELL. The Absorption Spectrum of Bismuth Oxide 44

Dr. S. RAIMES. Correlation Energy in Metals and the Cohesive Energy of Metallic Sodium 52

Mr. A. H. DE BORDE. The Auger Effect and Negative Meson Capture 57

Dr. R. F. BARROW and Dr. A. D. CAUNT. The Ultra-Violet Spectrum of a Diatomic Hydride Excited in Hydrogen–Potassium Fluoride Hollow-Cathode Discharges 68

Mr. R. WOLFE. On the Theory of Optical Absorption in Metals and Semiconductors 74

Mr. W. MARSHALL. Inelastic Magnetic Scattering of Neutrons from a Ferromagnetic Crystal 85

Letters to the Editor:

Mr. G. Z. SHAH, Mr. N. J. PATEL and Dr. K. M. GATHA. Scattering of High Energy Nucleons and Electrons by Carbon 92

Mr. P. G. DODSWORTH and Dr. R. F. BARROW. Triplet Electronic States of Aluminium Monofluoride 94

Mr. S. J. Q. ROBINSON and Dr. R. F. BARROW. The E–X Band System of SiS in Emission and the Dissociation Energy of SiS 95

Reviews of Books 97

Contents of Section B 100

Friction of an Elastic Solid *

By A. S. LODGE and H. G. HOWELL
The British Rayon Research Association, Urmston, near Manchester

MS. received 14*th October* 1953

Abstract. A study in terms of elastic asperity theory is made of the empirical frictional relation $F=\alpha R^n$ where F and R are frictional force and normal reaction respectively, α is a coefficient of friction and n a friction index. Three different contact cases are treated: (i) plane surfaces, (ii) cylindrical surfaces, (iii) a string round a cylinder. Expressions are derived for F (or the tension in case (iii)) in terms of R and the size, number and elastic constants of the asperities. This enables the frictional 'constants' to be related to these properties.

§ 1. Introduction

In the cohesion theory of friction, contact between two solid surfaces is considered to occur at the tips of irregularities or asperities, even when the bodies are apparently smooth. Since the area of such contacts is so small the pressures developed are sufficiently high to produce plastic flow of the materials. The consequent enlargement of the area of contact which occurs continues until the pressure has fallen to the yield value p_y at which point the area of contact is given by $a=r/p_y$ where r is the load supported by the individual asperity.

At this stage the surface materials are in such intimate contact that the asperities may for some materials become welded together and the tangential force f required to break such a contact can be represented as $f=Sa$, where S is the shear strength of the weld. This gives $f=Sr/p_y$ and since S and p_y are presumably constants for the system this results in the local friction law $f=\mu r$ with the constant $\mu=S/p_y$. If it be assumed that all such contacts break at the same moment under the application of an external shearing force F and all f's are parallel to F we have $F=\Sigma f=(S/p_y)\Sigma r=(S/p_y)R=\mu R$, where R is the total load.

This is Amontons's law of friction and it is seen that it has the same form as for the localized asperity. Further, the result is independent of the number and distribution of these asperities.

This simple approach accounts satisfactorily for most cases of metallic friction; only in the case of soft metals is a modification necessary. Here, under the combined application of the shear and normal forces, the contact area can grow to much bigger values than that given by the above argument (see McFarlane and Tabor (1950)).

In the non-metallic field many exceptions to Amontons's law have been reported (Bowden and Young 1951, Howell and Mazur 1953, Lincoln 1952, Schallamach 1952).

* This work forms part of a programme of fundamental research of the British Rayon Research Association.

In these cases the observed frictional force is related to the load by the empirical equation $F=\alpha R^n$, where α is a constant and n is usually found to lie between 2/3 and unity. An interpretation of the value $n=2/3$ can be given for a *single* contact of two spheres or of a sphere on a plane, or of two crossed cylinders if it be assumed that the contact area is now determined by Hertz's elastic theory (1881). This gives the area as proportional to $R^{2/3}$ and assuming as before that $F=S\times$ area we have $F\propto R^{2/3}$. As $n=1$ is the value obtained for plastic flow of a contact it is natural to suggest that values of n between 2/3 (ideally elastic) and 1 (purely plastic) mean that the area deformation is determined by visco-elastic properties. However, n is calculated from overall measurements on F and R and not on the forces involved in single asperities and it is not certain that the *observed* friction law is the same as that which operates at individual contacts as *is* the case when $F=\mu R$ on the cohesion theory (or $n=1$). Indeed Lincoln (1953) has reported that for the contact of two crossed cylinders with longitudinal fluted asperities a friction load index of 8/9 may be obtained for elastic deformation of the asperities. Further, in a recent paper Archard (1953) has pointed out the importance of any distribution in depth of the asperities

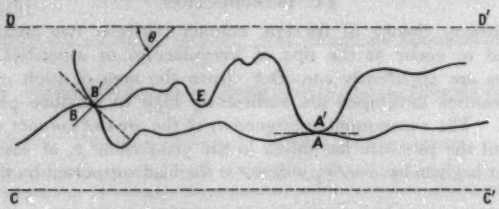


Figure 1. Contact of two ' plane ' surfaces.

(i.e. in a direction normal to contact surface) and has indicated that the index will be influenced by such a distribution. In view of this there arises the question as to the status of the 'coefficient of friction' in the case of contact between solids whose asperities are deformed elastically. If we regard Amontons's law $F=\mu R$ as a special form of $F=\alpha R^n$ we see that μ on the cohesion theory is a constant dependent only upon the properties of the surface materials whilst n is independent of these but depends only on the mechanism of deformation. In what light are we to regard these empirical constants in the case of purely elastic deformation? This paper attempts to answer this by developing from first principles the (F, R) relation of two such bodies in contact.

§2. Forces at a Single Contact

We consider two bodies making contact over a number of asperities as in figure 1. We use the word 'contact' to denote a separation of the surfaces within the range of molecular forces, i.e. say within 10 Å. The following assumptions will be made.

(i) Tangent planes to all contact regions are parallel to the main, average, surface directions nearby. In the case of planes in contact all tangent planes are parallel to these planes—such a case is at the contact AA′ in the diagram

where the tangent plane is parallel to CC′ and DD′. In practice the tangent planes at the asperities will not all be parallel to the main surface direction but will be distributed in some way about this direction. If θ denotes the angle which the normal to a contact region makes with the normal to one of the main surfaces as at BB′, this distribution can be described by a function $n(\theta)$, where $n(\theta)\,d\theta$ gives the fraction of contact regions whose angles θ lie in the range $(\theta, \theta+d\theta)$; $n(\theta)$ will presumably be a symmetric distribution about the mean $\theta=90°$. Assumption (i) is that this distribution is sufficiently sharp for us to neglect contributions to the total frictional force arising from contacts with values away from this mean.

(ii) The deformation is purely elastic at each contact and governed by a modified form of the Hertz equation.

(iii) The application of the frictional force does not affect the magnitude of the contact area as determined from the Hertz theory. Justification for this assumption is given by Mindlin's (1949) recent extension of Hertz's calculations to allow for the presence of external forces applied parallel to the contact region (Hertz's theory allows for normal forces only); Mindlin finds that the contact region can at most suffer rigid displacement, without change of size or shape, when tangential forces are applied to the bodies in contact.

(iv) At each contact the same local friction law is obeyed, and is of the form

$$f=\alpha r^n, \qquad \ldots\ldots(1)$$

where f and r are the limiting friction force and normal reaction respectively, both referring to unit area of contact. This is somewhat more general than the usual treatment, which assumes $f=S\times$ area, with S independent of r, in which case $n=0$. This is satisfactory in the case of plastic flow, but it may be that in the elastic case as the pressure increases and approaches p_y, changes in contact separation occur which increase the cohesive forces and hence also S, which thus becomes a function of r. At least it seems reasonable to allow for such a possibility by expressing the friction law in the above form.

(v) Friction becomes limiting at all contact regions simultaneously.

(vi) All contact regions are circular.

2.1. *Forces at a Typical Contact Region*

Let

$$F_i=\iint_{A_i} f\,dA \text{ and } R_i=\iint_{A_i} p\,dA \qquad \ldots\ldots(2)$$

denote the total limiting friction and normal reaction or load at the ith contact, A_i being the total area of the ith contact.

From Hertz's theory, the dependence of the area A_i on R_i is known and is of the form

$$A_i=k_i R_i^{2/3}, \qquad \ldots\ldots(3)$$

where k_i depends on the elastic constants of the bulk material, which are the same for all contacts, and upon the curvature of the ith contact before deformation. Now this theory assumes a small area of contact (i.e. when compared with the dimensions of the asperity) and this is probably far from true in the present case. If this restriction is discarded it seems physically plausible to use a relation of the same form, viz.

$$A_i=k_i R_i^{\beta} \qquad \ldots\ldots(4)$$

but with some value of β other than 2/3 which, we suppose, will be the same for

all contact regions, i.e. is independent of i; k_i here will be different from k_i in (3) but can still be regarded as a constant.

From (1), (2) and (4) we have

$$F_i = \iint_{A_i} \alpha p^n dA, \quad R_i = \iint_{A_i} p\, dA. \qquad \ldots\ldots(5)$$

To proceed further we need to know just how p varies over the contact region and this depends on the shape of the contact region—circular, elliptical or irregular. We consider them to be all circular with a circularly symmetric stress distribution. As in eqn (4) this may not be exactly as given by Hertz but it is reasonable to expect it to be of the same form and so we write it as

$$p \propto (1 - x^2/a_i^2)^h, \qquad \ldots\ldots(6)$$

where h is the same for all contacts and a_i is the radius of the ith contact region. The Hertz case is the particular one with $h = \frac{1}{2}$. Expressing the constant of proportionality in (6) in terms of the total reaction R_i we obtain

$$p = \frac{h+1}{\pi a_i^2} R_i (1 - x^2/a_i^2)^h, \qquad \ldots\ldots(7)$$

which gives the normal stress at a distance x from the centre of the asperity.

When this expression is substituted for p in (5), we find

$$F_i = \int_{\phi=0}^{2\pi} d\phi \int_{x=0}^{a_i} \alpha \left[\frac{h+1}{\pi a_i^2} R_i \left(1 - \frac{x^2}{a_i^2}\right)^h\right]^n x\, dx = \frac{\alpha \pi a_i^2}{1+hn} \left[\frac{h+1}{\pi a_i^2}\right]^n R_i^n. \qquad \ldots(8)$$

Now πa_i^2 can be expressed in terms of R_i by using (4), viz. $k_i R_i^\beta = A_i = \pi a_i^2$. Thus

$$F_i = \alpha \frac{(h+1)^n}{1+hn} k_i^{1-n} R_i^{\beta(1-n)+n}, \qquad \ldots\ldots(9)$$

which is the relation between the total frictional force and the total normal reaction at the ith contact.

§3. Resultant Force over all Contacts

The total frictional force and the total normal load F_T, R_T (i.e. the observables in an actual friction experiment) are given by the equations

$$F_T = \Sigma F_i, \quad R_T = \Sigma R_i. \qquad \ldots\ldots(10)$$

Hence the relation between F_T and R_T is contained in the equations

$$F_T = \alpha \sum_i \frac{(h+1)^n}{1+hn} k_i^{1-n} R_i^{\beta(1-n)+n}, \quad R_T = \sum_i R_i, \qquad \ldots\ldots(11)$$

and more explicit expressions between them can only be obtained by considering certain cases.

3.1. *Contact of Plane Surfaces*

We assume that all the asperities are identical in nature, i.e. they all have the same curvature before contact and bear the same load R_i. This requires the k_i to be equal and a uniform distribution of the external load.

Let the total number of contacts be N.

Equation (11) now becomes

$$F_T = N\alpha \sum_i \frac{(h+1)^n}{1+hn} k_i^{1-n} R_i^{\beta(1-n)+n}, \quad R_T = N R_i, \qquad \ldots\ldots(12)$$

and, eliminating R_i we get

$$F_T = \alpha_T R_T^{n_T} \qquad \ldots\ldots(13)$$

where $$\alpha_T = \alpha N^{(1-n)(1-\beta)} \frac{(h+1)^n}{1+hn} k^{1-n}, \qquad \ldots\ldots(14)$$

k being the common value of k_i,

and $$n_T = \beta(1-n)+n. \qquad \ldots\ldots(15)$$

We observe (i) that the observed frictional force F_T depends upon the number N and geometry k of the contacts, i.e. upon the state of roughness of the surface, and (ii) that the observed index n_T is only equal to the force-area index β of the asperity when $n=0$, i.e. when there is no load dependence of S, the shear strength of the contacts, or when $\beta=1$, which is essentially the condition for plastic deformation and will not be considered here further. This treatment ignores any distribution in depth of the asperities. In practice asperities such as E in the diagram will assume some share of the load at some higher value of R_T. Contacts of this type will destroy the uniformity of the load distribution and this will give rise to a more complex relation between F_T and R_T depending upon the particular distribution assumed. It will be observed, however, that as N is now some monotonic increasing function of R_T, the inclusion of this function in equations of type (13) or (14) will effectively give a bigger index n_T than that indicated by eqn (15).

3.2. *Contact of Curved Surfaces*

Another case in which N will vary with the applied load is that of the contact of two curved surfaces, such as two spheres, or two crossed cylinders. Here, as the load is increased, more and more outlying areas are brought into contact.

Here we consider surfaces in which the asperities are all alike and sufficiently numerous for summations over the contacts to be approximated to by integrals.

Let the surface density of asperities be σ.

The total load R_i on an asperity will depend upon its position in the overall contact region (which is circular) according to the equation

$$p=(1-x^2/a^2)^{1/2}\, 3R_T/2\pi a^2 \qquad \ldots\ldots(16)$$

where a is the radius of the overall contact region and p the pressure at radia distance x from the centre of the region.

Consider an annulus of this contact region bounded by circles of radii x, and $x+dx$ concentric with the boundary $x=a$. The force on this annulus given by the Hertz theory must be balanced by the sum of the loads on the asperities in the annulus; hence

$$p(x)\, d(\pi x^2) = \sum_{\substack{\text{all contacts in} \\ \text{the annulus}}} R_i. \qquad \ldots\ldots(17)$$

Taking the annulus so narrow that the variation of R_i for these contacts may be neglected we have

$$p(x)\, d(\pi x)^2 = R_i \times \text{no. of contacts} = R_i \sigma\, d(\pi x^2).$$

Hence (16) becomes

$$\sigma R_i = (1-x_i^2/a^2)^{1/2}\, 3R_T/2\pi a^2, \qquad \ldots\ldots(18)$$

putting $x=x_i$, which shows how the load R_i on an asperity depends on its distance x_i from the centre of the overall contact region. Now the total tangential force across the contact surface is given by

$$F_T = \frac{\alpha(1+h)^n}{1+hn} \sum_{\text{all contacts}} k_i^{1-n} R_i^{2\gamma}, \text{ writing } \beta(1-n)+n=2\gamma.$$

Substituting for R_i from eqn (18) and using the assumption that all k_i are the same, $=k_0$ say, we have

$$F_T = \frac{\alpha(h+1)^n}{1+hn} k_0^{1-n} \frac{3}{2\pi}\left[\frac{R_T}{\sigma\alpha^2}\right]^{2\gamma} \sum_i \left(1-\frac{x_i^2}{a^2}\right)^\gamma. \qquad \ldots\ldots(19)$$

To evaluate the summation we note that in the annulus, x, $x+dx$ there are $\sigma\, d(\pi x^2)$ contacts, and so

$$\sum_i \left(1-\frac{x_i^2}{a^2}\right)^\gamma = \int_{x=0}^{a} \left(1-\frac{x^2}{a^2}\right)^\gamma d(\pi x^2) = \frac{\sigma\pi a^2}{\gamma+1}.$$

Hence

$$F_T = \alpha \frac{(h+1)^n}{1+hn} k_0^{1-n}\left[\frac{3R_T}{2\pi\sigma a^2}\right]^{2\gamma} \frac{\sigma\pi a^2}{\gamma+1}. \qquad \ldots\ldots(20)$$

Now $\pi a^2 = A = \kappa R_T^{2/3}$, κ depending upon the dimensions and elastic constants of the spheres or cylinders, and so

$$F_T = \frac{2(\frac{3}{2})^{\beta(1-n)+n}}{1+\beta(1-n)+n} \frac{\alpha(h+1)^n}{1+hn} k_0^{1-n} (\kappa\sigma)^{(1-n)(1-\beta)} R_T^{\{2+\beta(1-n)+n\}/3}$$

$$= \text{constant} \times R_T^{\{2+\beta(1-n)+n\}/3}. \qquad \ldots\ldots(21)$$

In the case of two equal spheres or cylinders, of radius ρ, and $\beta = 2/3$, $n=0$,

$$F_T \propto R_T^{8/9} \sigma^{1/3} \rho^{2/9},$$

so that in this case a friction index of 8/9 can be obtained even when the asperities are deformed elastically.

3.3. *Contact of a String round a Cylinder*

One of us (Howell 1953) has recently shown that for limiting friction, the initial and final tensions T_0, T of a string wrapped round a cylinder of radius ρ through an angle θ is given by the equation

$$T^{1-n} = T_0^{1-n} + (1-n)\alpha\theta\rho^{1-n} \qquad \ldots\ldots(22)$$

where the assumed friction law is $F = \alpha R^n$ and the contact of the string is continuous. This section is devoted to the re-examination of this equation on the asperity model.

The problem involves the determination of the apparent contact area of an elastic string round a rigid cylinder, and we attempt a solution by treating it as one of the string in contact with a plane, which appears reasonable if the radius of curvature of the cylinder is much larger than that of the string. This enables the use to be made of the Hertz relation between the width $2b$ of the contact region and the normal force per unit length P, viz.

$$b = \left\{\frac{1}{\pi}\left[\frac{1-\sigma_1^2}{E_1} + \frac{1-\sigma_2^2}{E_2}\right]\rho_s P\right\}^{1/2} \qquad \ldots\ldots(23)$$

where ρ_s is the radius of the cylindrical body (the string) and σ_1, σ_2, E_1, E_2 are moduli for cylinder and plane.

Referring to figure 2 which shows the overall system (a) and an element of the system (b) we have for the equilibrium of the latter

$$dR = T d\theta, \quad dT = dF. \qquad \ldots\ldots(24)$$

Also

$$dR = P\rho_c\, d\theta \ (\rho_c = \text{cylinder radius}) \qquad \ldots\ldots(25)$$

so that

$$P\rho_c = T. \qquad \ldots\ldots(26)$$

Now as we move along the string in the direction of increasing tension T, P will increase, from (26), and so also will b, from (23). Therefore if the region

of apparent contact could be straightened out, it would look something like figure 2 (*c*).

The normal force over this area will vary laterally in a manner given by Hertz, and also longitudinally.

Considering first the lateral variation, the normal stress component p varies with y according to

$$p = \frac{2P}{\pi b}\left(1 - \frac{y^2}{b^2}\right)^{1/2}. \qquad \ldots\ldots(27)$$

Now the number of asperities in the element $dx\,dy$ shown is $\sigma\,dx\,dy$, and if we take this element so small that the variation of p over it is negligible,

$$p = \sigma R_i, \qquad \ldots\ldots(28)$$

where R_i, as before, is the normal force on the ith asperity. We can relate R_i to

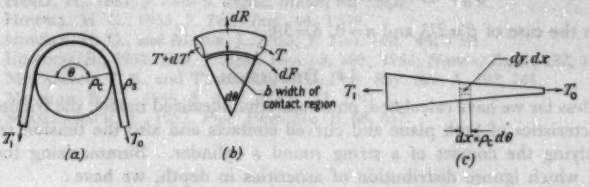


Figure 2. Contact of string round cylinder.

the frictional force at that asperity from eqn (9), and we must also evaluate the dF of (24) which will be

$$dF = \sum_{\substack{\text{all contacts in} \\ \text{the strip } x,\, x+dx}} F_i \qquad \ldots\ldots(29)$$

and to do this we replace it by an integral over y, relating R_i to p. Thus

$$dF = \frac{\alpha(h+1)^n}{1+hn}\, k_0^{1-n} \sum R_i^{2\gamma}$$

$$= \frac{\alpha(h+1)^n}{1+hn}\, k_0^{1-n} \int_{y=-b}^{b} \left(\frac{p}{\sigma}\right)^{2\gamma} \sigma\,dx\,dy \qquad \ldots\ldots(30)$$

(since there are $\sigma\,dx\,dy$ contacts in the element)

$$= A \int_{-b}^{b} \left[\frac{2P}{\pi b \sigma}\right]^{2\gamma} \left(1 - \frac{y^2}{b^2}\right)^{\gamma} \sigma\,dx\,dy \qquad \ldots\ldots(31)$$

where $A = \dfrac{\alpha(h+1)^n}{1+hn}\, k_0^{1-n}$,

$$= A\left[\frac{2P}{\pi b \sigma}\right]^{2\gamma} B(\tfrac{1}{2}, 1+\gamma) b\sigma\,dx \qquad \ldots\ldots(32)$$

where $B(\frac{1}{2}, 1+\gamma)$ is the beta function defined by the equation

$$B(m, n) = \int_0^1 q^{m-1}(1-q)^{n-1}\,dq \quad (m>0,\ n>0).$$

Here $q = y^2/b^2$.

From (24),

$$dT = A(2P/\pi)^{2\gamma} B(\tfrac{1}{2}, 1+\gamma) b\sigma\rho_c\,d\theta.$$

Eliminating b from (23),

$$dT = A\left(\frac{2P}{\pi}\right)^{2\gamma} B(\tfrac{1}{2}, 1+\gamma)\left\{\frac{\sigma^2}{\pi}(\kappa_1+\kappa_2)\rho_c P\right\}^{\frac{1}{2}-\gamma} \rho_c\,d\theta, \qquad \ldots\ldots(33)$$

$[\kappa_1 = (1-\sigma_1^2)/E_1]$ etc., and finally from (26) we can express P in terms of T:

$$dT = A\left(\frac{2}{\pi}\right)^{2\gamma} B(\tfrac{1}{2}, 1+\gamma)\left\{\frac{\sigma^2}{\pi}(\kappa_1+\kappa_2)\rho_s\right\}^{\frac{1}{2}-\gamma} \rho_c\left(\frac{T}{\rho_c}\right)^{\frac{1}{2}+\gamma} d\theta. \quad \ldots\ldots(34)$$

Writing

$$\delta = \gamma + \tfrac{1}{2} = \tfrac{1}{2}(1-n)(1+\beta)+n,$$

and integrating, we have, provided $\delta \neq 1$,

$$\frac{T^{1-\delta} - T_0^{1-\delta}}{1-\delta} = M\theta\rho_c^{1-\delta}, \quad \ldots\ldots(35)$$

where

$$M = \left(\frac{2}{\pi}\right)^{\beta(1-n)+n} B(\tfrac{1}{2}, 1+\gamma)\,\frac{a(h+1)^n}{1+hn}\, k_0^{1-n}\left\{\frac{1}{\pi}\left[\frac{1-\sigma_1^2}{E_1} + \frac{1-\sigma_2^2}{E_2}\right]\right\}^{(1-n)(1-\beta)/2}$$
$$\times\, \sigma^{(1-n)(1-\beta)}\rho_s^{(1-n)(1-\beta)/2}. \quad \ldots\ldots(36)$$

In the case of $\beta = 2/3$ and $n = 0$, $\delta = 5/6$.

§4. Discussion

Thus far we have calculated, on a somewhat idealized model, the friction–load characteristics of both plane and curved contacts and also the tension relations underlying the contact of a string round a cylinder. Summarizing for these cases, which ignore distribution of asperities in depth, we have:
for the planes

$$F = \alpha_P R^{n_P}, \quad n_P = 2\gamma = \beta + n(1-\beta), \quad = \beta \text{ if } n = 0,$$

for curved surfaces

$$F = \alpha_C R^{n_C}, \quad n_C = \tfrac{2}{3}(1+\gamma) = \tfrac{1}{3}\{\beta + n(1-\beta)\} + \tfrac{2}{3}, \quad = (\beta+2)/3 \text{ if } n = 0,$$

and for the string round the cylinder

$$T^{1-\delta} = T_0^{1-\delta} + (1-\delta)M\theta\rho_c^{\delta}, \quad \delta = \tfrac{1}{2} + \gamma = \tfrac{1}{2}(1-n)(1+\beta)+n, \quad = (\beta+1)/2 \text{ if } n = 0.$$

We have here three pairs of constants consisting of a coefficient α_P, α_C, or M, and a corresponding index n_P, n_C, or δ. Of these constants, the index is independent of the nature of the contacting materials (if there is no dependence of shear properties on the load) but does depend upon the relative geometrical configuration of the surfaces. Thus in the simplest case of two planes in contact, it coincides with the deformation index β and this is modified if the type of contact is changed to either the two cylinders case or the string round the cylinder case. The coefficient α, however, differs from the classical coefficient μ in that it is not merely a function of the mechanical properties of the surface material but involves the size, shape, distribution and concentration of the contacting asperities. This clearly shows the difficulty of interpreting the coefficient in terms of any material constants such as can be done with the classical coefficient.

It will be further noted that this treatment indicates that a friction–load relation as determined by a particular experiment cannot be unique but is dependent upon the particular geometrical arrangement of the contactin g surfaces, i.e. ultimately it will depend upon the particular experimental method.

These cases demonstrate that the measured friction index can be expected to be greater than 2/3 even though perfectly elastic deformation of the asperities is assumed. With plane surfaces one can anticipate the nearest approach to this figure: it should be 2/3 plus the increment caused by the distribution in depth.

Schallamach's value of 2/3 with flat rubber surfaces is therefore in accordance with this expectation. When we consider real surfaces in practice, we must be prepared for departures from the idealized model treated here. In particular the effect of the distribution in depth of the asperities may be expected to smooth out the differences in n_P, n_C, δ so that it may be difficult to detect such changes experimentally particularly in view of the usual large experimental scatter in friction work. Or, alternatively, it may well be that the experimental scatter is a consequence of localized changes in geometry of the contacting surfaces.

References

Archard, J. F., 1953, *J. Appl. Phys.*, **24**, 981.
Bowden, F. P., and Young, J. E., 1951, *Proc. Roy. Soc.* A, **208**, 444.
Hertz, H., 1881, *J. reine u. angew. Math.*, **92**, 156.
Howell, H. G., 1953, *J. Text. Inst.*, **44**, T359.
Howell, H. G., and Mazur, J., 1953, *J. Text. Inst.*, **44**, T59.
Lincoln, B., 1952, *Brit. J. Appl. Phys.*, **3**, 260; 1953, *Nature, Lond.*, **172**, 169.
McFarlane, J. S., and Tabor, D., 1950, *Proc. Roy. Soc.* A, **202**, 244.
Mindlin, R. D., 1949, *J. Appl. Mech.*, **71**, 259.
Schallamach, A., 1952, *Proc. Phys. Soc.* B, **65**, 657.

Discontinuous Flow in Zinc Crystals and its Relationship to Strain Ageing

By M. J. DUMBLETON*
Department of Metallurgy, University of Birmingham

Communicated by A. H. Cottrell; MS. received 12th October 1953

Abstract. It was found that nitrogen-treated zinc crystals could strain-age under an applied stress, and also during creep provided the creep rate was small enough. A repeated yielding, strongly dependent on temperature and rate of strain, and which seemed to be due to strain ageing during flow, was observed in stress-strain tests. Jerky flow in stress-strain tests, and jerky stress-relaxation curves were observed over a much wider range of conditions, and appeared also to be due to the presence of nitrogen acting in some manner other than strain ageing.

§ 1. Introduction

Strain ageing during plastic flow is well known from the repeated yielding which occurs when iron is strained at temperatures in the blue brittleness range (Fettweiss 1919, Hall 1952, Paxton and Churchman 1953). Similar phenomena have been observed in other metals and alloys, in the form of serrations in the stress-strain curves (Rosenhain and Archbutt 1912, Portevin and Le Chatelier 1923, Anderson 1926, Schmid and Valouch 1932, Fell 1936, Kuroda 1938, McReynolds 1949, Lubahn 1949, Kawada 1950, 1952, Berghezan 1952, Ardley and Cottrell 1953) or as jerky flow in creep experiments (Becker and Orowan 1932, Ardley and Cottrell 1953), and have also been observed in rock salt (Joffé 1928, Classen-Nekludowa 1929), but it is only in the case of iron and soft steel that a connection with strain ageing has been clearly established.

However, it has been shown recently that strain ageing can occur in single crystals of zinc, if these contain a small amount of, for example, nitrogen (Wain and Cottrell 1950, Dumbleton and Howlett 1952, Wain 1952). In the experimental work described below the possibility of strain ageing during flow in zinc has been investigated. The conditions for discontinuous flow have been examined, and its relationship to strain ageing considered.

Single crystals were grown by the fusion method (Andrade and Roscoe 1937) in the form of wires of 1 mm diameter, and these were electropolished after mounting between copper wire grips. Two grades of zinc were used, 'pure' and 'commercial', the spectrographic analyses of which are as follows:

	Pb (%)	Cd (%)	Cu (%)	Fe (%)	Purity (%)
Pure	0·001	0·0001	0·0005	0·001	99·997
Commercial	0·010	0·010	0·003	0·003	99·97

The techniques of impregnating the pure zinc with nitrogen, of electropolishing, and of handling the crystals generally, were identical with those used and described by Wain and Cottrell (1950).

* Now at Department of Physics, University of Birmingham.

Two kinds of experiments were made on these crystals. In §2 an account is given of the creep experiments, which showed that, if the rate of creep becomes sufficiently slow, strain ageing can occur and prevent further flow until the applied stress is increased substantially. Later experiments, described in §3, were made on a fairly hard testing machine of the Polanyi type and showed that, under suitable conditions, yielding and strain ageing could occur repeatedly, producing an effect analogous to the blue brittleness of iron. Another type of serration of the stress-strain curve, also depending on the presence of impurity atoms, was observed over a much wider range of conditions. Jerky flow in stress relaxation experiments was also observed. The results are summarized and discussed in §4.

§2. Creep Experiments on Pure Zinc Crystals containing Nitrogen

These experiments were made using the creep apparatus described by Cottrell and Aytekin (1950), in which a template enables constant resolved shear stress to be maintained on the active slip system in the crystal. A bath of

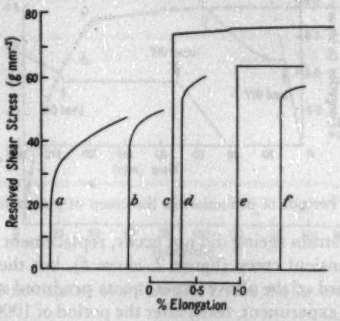


Figure 1. Stress-strain curves of pure zinc crystals. *a*, without nitrogen, 20°C; *b*, immediate re-test; *c*, with nitrogen, 20°C (strain-aged at 165°C); *d*, immediate re-test; *e*, with nitrogen, 50°C (strain-aged under stress of 50·7 g mm^{-2} at 50°C); *f*, immediate re-test.

silicone oil was used to keep the temperature of the specimen constant during an experiment. The apparatus could be adapted for making stress-strain tests, without moving the specimen, by loading water from a burette into a light cup suspended from the loading pan. This machine was of the 'dead loading' type, so that when a sharp yield occurred there was no drop in load but a sudden onset of rapid deformation, causing the stress-strain curve to change sharply from the linear elastic region to a nearly horizontal plastic region (figure 1, curve *c*). If the specimen was then unloaded before much flow had occurred, a second loading applied immediately showed that in this overstrained specimen plastic flow set in smoothly at a stress lower than the initial yield (figure 1, curve *d*). Crystals of the pure zinc behaved in this way, of course, only when they had been treated with nitrogen; otherwise they behaved as in figure 1, curves *a* and *b*.

In the attempt to produce strain ageing during creep advantage was taken of the fact that a sudden small reduction of stress on a specimen flowing at a constant rate produces a 'period of induction', during which the rate of flow is very small or zero, before the new steady rate sets in (Cottrell and Aytekin). Figure 2, curve *a* shows an example of this caused by reducing the resolved shear stress from 59·6 g mm^{-2} to 53·9 g mm^{-2} on a nitrogen-bearing pure zinc crystal at 20°c. The period of induction was about 12 minutes. No strain ageing occurred, or was expected, since for this material at this temperature strain ageing takes about 8 hours (Wain and Cottrell 1950).

In similar experiments on another specimen the temperature was held at 50°c, the expected time of strain ageing then being about fifteen minutes. The effects of reducing the applied stress by various amounts, starting always from a state of steady flow under a resolved shear stress of 54·8 g mm^{-2}, were as follows:

Reduction of stress (g mm^{-2})	2·0	3·1	4·1
Induction time (min)	1	5	No resumption of flow in 1000 min

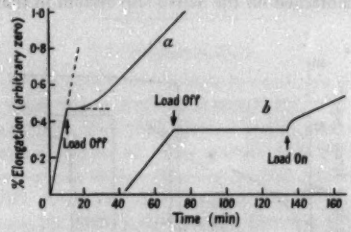


Figure 2. Periods of induction in the creep of pure zinc crystals.

In those cases where strain ageing did not occur, replacement of the load removed produced a rapid transient creep (figure 2, curve *b*), but the replacement of the 4·1 g mm^{-2} in the third of the above experiments produced no plastic flow at all. In fact a stress–strain experiment, taken after the period of 1000 minutes, produced no plastic flow until the stress was raised to 65 g mm^{-2}, at which value a sudden yield took place (figure 1, curves *e* and *f*). Thus strain ageing had occurred.

These experiments raised the possibility that strain ageing might be made to occur, not only during a period of induction, but also while the material was actually flowing at a perceptible rate. For even during continuous plastic flow individual dislocations should move discontinuously, halting at obstacles and, when an obstacle is overcome, gliding forward rapidly to the next one. If a dislocation were made to halt at an obstacle for a sufficiently long time it might become anchored by solute atoms and not move again. To examine this, creep experiments were made at temperatures about 100°c, where strain ageing occurs within a few seconds, under a continuously decreasing stress. The latter was applied by means of a soft spring which relaxed as the specimen extended. Figure 3 shows a curve obtained in this way at 90°c. At the point P a sharp discontinuity occurred in the rate of flow and, after a small further extension ($1{\cdot}7 \times 10^{-3}$ cm on a 5 cm gauge length) to the point Q, flow ceased completely. A subsequent exploration in a stress-strain experiment showed that the specimen had become strain aged.

§ 3. Experiments with a Hard Tensile Machine

A machine of the Polanyi (1925) type was used. The stress was measured by means of an optical lever from the deflection of an elastic beam in the loading circuit, and a constant cross-head speed was obtained from a drive geared to a gramophone motor. A drum camera enabled stress-strain curves to be recorded photographically. In such a machine the sudden extension of a specimen on yielding causes the load on it to fall to the value where the rate of elongation equals the cross-head speed. Stress-relaxation experiments can also be made; here the specimen is stressed quickly to a certain initial value, the cross-head is then halted and the course of the subsequent extension (or stress-relaxation, since this is coupled to the extension) is recorded.

Yield points were developed in crystals of commercially pure zinc by the usual straining and ageing treatment. Typical stress-strain curves for this material are shown in figure 4. At -18°c the curve is smooth, once the initial yield has occurred, but at higher temperatures the flow takes place jerkily. The serrations become sharper as the temperature is raised, because the load

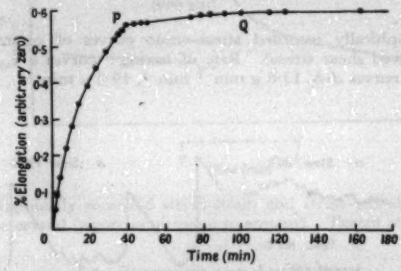


Figure 3. Strain ageing of a pure zinc crystal during creep at 90°c.

falls more quickly at each jerk. However, the size of the jerks does not vary much over a wide range of temperature; nor did it vary much with the rate of application of stress. This is in contrast with the behaviour expected of an effect depending on successive yielding and strain ageing, such as blue brittleness in iron. Behaviour of the latter type was, however, observed under some conditions, as shown in the curves of figure 5. In curve *a*, which was obtained at 80°c and a rate of stressing of 2·1 g mm^{-2} min^{-1}, the jerks are similar to those already described. But when the temperature was raised to 105°c large jerks were developed (curve *b*). When, after a short rest, the rate of stressing at this same temperature was increased to three times its former value (curve *c*), the jerks were reduced to only one fifth of their previous size, even though the initial yield was as large as the repeated yields which constitute the jerks in curve *b*. We shall distinguish this behaviour from the more widely observed jerky flow, described earlier, by calling it 'repeated yielding'.

The stress-relaxation curves shown in figure 6 were obtained by first straining rapidly, to reach a high initial stress, and then observing the fall in stress while the cross-head remained fixed. The commercial zinc crystals gave stepped

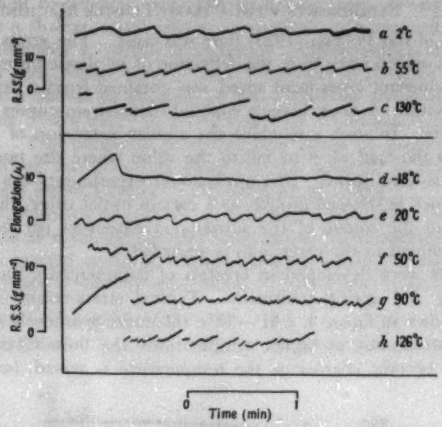


Figure 4. Photographically recorded stress-strain curves of commercial zinc crystals (R.S.S. = resolved shear stress). Rate of testing: curves *a–c*, 9·5 g mm^{-2} min^{-1}, 6·1 μ min^{-1}; curves *d–h*, 13·6 g mm^{-2} min^{-1}, 19·2 μ min^{-1}.

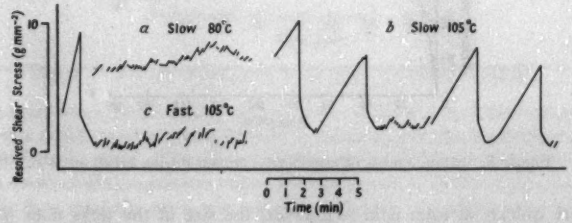


Figure 5. Repeated yielding in a commercial zinc crystal. Rate of testing: slow, 2·1 g mm^{-2} min^{-1}, 6·5 μ min^{-1}; fast, 6·3 g mm^{-2} min^{-1}, 18 μ min^{-1}.

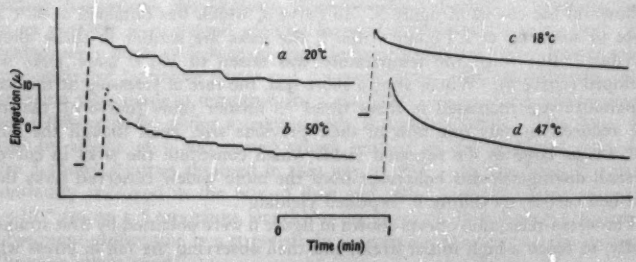


Figure 6. Photographically recorded stress-relaxation curves of zinc crystals. Curves *a* and *b*, commercial zinc; curves *c* and *d*, pure zinc.

curves (*a* and *b*), the sudden falls in stress being smaller but more frequent at 50°c than at room temperature.

Stress-strain and stress-relaxation curves on pure zinc crystals were smooth, showing neither sharp yielding nor jerky flow (figure 6, curves *c* and *d*). Pure zinc crystals treated with nitrogen showed an initial yield, but the subsequent curves were usually smooth, although sometimes slightly irregular. However, jerky flow and stepped relaxation curves could be produced in such crystals if the pure zinc from which they were grown had first been given an unusually vigorous nitrogen treatment (figure 7). This consisted of keeping the surface of the molten metal free from oxide for thirty minutes by liberal sprinklings of zinc chloride and ammonium chloride, and casting into a glass mould and cooling rapidly to retain as much gas as possible in the solid metal. It seems that jerky flow and stepped relaxation curves can be caused by nitrogen, but that more nitrogen is needed for these phenomena than for the sharp initial yield.

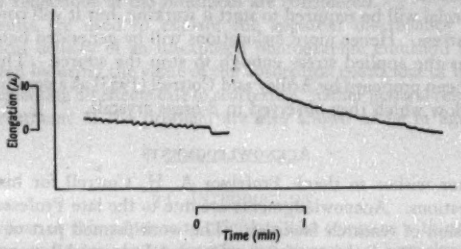


Figure 7. Photographically recorded stress-strain and stress-relaxation curves of a pure zinc crystal (rigorous nitrogen treatment). Tested at 50°c.

§ 4. Summary and Discussion

The stepped relaxation appears to be the same phenomenon as that observed by Becker and Orowan (1932). The zinc used for their single crystals was 'Kahlbaum 1928' which, from the analyses of Kahlbaum zincs of neighbouring years reported by Schmid and Valouch (1932), appears very similar to the commercial zinc used in the present investigation. Schmid and Valouch observed a jerky stress-strain curve in zinc of a purity similar to our pure zinc; their crystals were grown by the Czochralski method of withdrawal from the melt, zinc chloride being used as a flux, and the molten metal may have been exposed to nitrogen or air.

The effect of nitrogen on the plastic flow of zinc can be summarized as follows. A small nitrogen content produces a sharp yield point, followed by smooth flow. At the same time, the stress-relaxation curve is smooth. In creep experiments strain ageing can take place during flow, provided the creep rate is small. With greater nitrogen contents the stress-relaxation curve becomes stepped, and the stress-strain curve is jerky. At around 100°c repeated yielding is observed at slow rates of stressing. This repeated yielding disappears when the temperature is reduced or the rate of stressing increased, although the jerky flow remains. The characteristics of this repeated yielding suggest that it is caused by strain ageing during flow, similar to the blue brittleness phenomenon in iron. In contrast, the jerky flow does not vary markedly with temperature

and rate of stressing, and it is difficult to see how the steps of the stress-relaxation curves, which are a manifestation of jerky flow, could be explained by strain ageing during flow.

Mott (1952), and Fisher, Hart and Pry (1952), have suggested a way in which a Frank–Read source might give rise to a burst of slip on a slip plane. They point out that the stress required to generate the first ring of dislocation is greater than for the subsequent ones because of the kinetic energy which the generating dislocation acquires. Fisher, Hart and Pry calculate that about 300 dislocations should be generated before the back stress which they exert on the source counteracts this effect and causes the source to stop working. Jerks of this size would not be detected by the apparatus used in our experiments. In jerky flow the larger jerks observed corresponded typically to the passage across the crystal of 5000–10 000 dislocations. Remembering that detectable jerks occur only in nitrogen-bearing specimens, we can suggest an explanation. If the dislocation line of a Frank–Read source is locked by solute atoms, a greater stress than in the pure material will be required to start it working, but it will continue to work at the usual stress. Hence more dislocations will be generated before their back stress reduces the applied stress enough to stop the source. This mechanism has recently been proposed by Ardley and Cottrell (1953) as a possible explanation of the jerky flow which they observed in β-brass crystals.

Acknowledgments

The author wishes to thank Professor A. H. Cottrell for his interest and helpful suggestions. Acknowledgments are due to the late Professor D. Hanson for the provision of research facilities. The work formed part of a programme supported by the Department of Scientific and Industrial Research through the National Physical Laboratory.

References

Anderson, R. J., 1926, *Amer. Soc. Test. Mat., Proc.* **26,** part II (Technical papers), 349.
Andrade, E. N. da C., and Roscoe, R., 1937, *Proc. Phys. Soc.*, **49,** 152.
Ardley, G. W., and Cottrell, A. H., 1953, *Proc. Roy. Soc.* A, **219,** 328.
Becker, R., and Orowan, E., 1932, *Z. Phys.*, **79,** 566.
Berghezan, A., 1952, *Rev. Met.*, **49,** 99.
Classen-Nekludowa, M., 1929, *Z. Phys.*, **55,** 555.
Cottrell, A. H., and Aytekin, V., 1950, *J. Inst. Met.*, **77,** 389.
Dumbleton, M. J., and Howlett, B. W., 1952, *Proc. Phys. Soc.* B, **65,** 882.
Fell, E. W., 1936, *Iron St. Inst. Carnegie Scholarship Mem.*, **26,** 123.
Fettweiss, F., 1919, *Stahl u. Eisen*, **39,** 1, 34.
Fisher, J. C., Hart, E. W., and Pry, R. H., 1952, *Phys. Rev.*, **87,** 958.
Hall, E. O., 1952, *J. Iron St. Inst.*, **170,** 331.
Joffé, A. F., 1928, *The Physics of Crystals*, (Ed. L. B. Loeb) (London : McGraw–Hill), p. 50.
Kawada, T., 1950, *J. Phys. Soc. Japan*, **5,** 60, 64, 68; 1952, *Ibid.*, **7,** 240.
Kuroda, M., 1938, *Sci. Papers, Inst. Phys. Chem. Res., Tokyo*, **34,** 1528.
Lubahn, J. D., 1949, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **185,** 702.
McReynolds, A. W., 1949, *Trans. Amer. Inst. Min.* (*Metall.*), *Engrs*, **185,** 32.
Mott, N. F., 1952, *Phil. Mag.*, **43,** 1151.
Paxton, H. W., and Churchman, A. T., 1953, *Acta Metallurg.*, 473.
Polanyi, M., 1925, *Z. Tech. Phys.*, **6,** 121.
Portevin, A., and Le Chatelier, F., 1923, *C. R. Acad. Sci., Paris*, **176,** 507.
Rosenhain, W., and Archbutt, S. L., 1912, *Proc. Instn Mech. Engrs*, Part 2, 319.
Schmid, E., and Valouch, M. A., 1932, *Z. Phys.*, **75,** 531.
Wain, H. L., 1952, *Proc. Phys. Soc.* B, **65,** 886.
Wain, H. L., and Cottrell, A. H., 1950, *Proc. Phys. Soc.* B, **63,** 339.

The Application of Radiative Transfer Theory to Scattering Effects in Unexposed Photographic Emulsions*

By E. PITTS

Research Laboratories, Kodak Ltd., Wealdstone, Harrow, Middlesex

MS. received 8th June 1953, and in amended form 23rd November 1953

Abstract. The general problem of the transmission of light through a medium containing many particles which both scatter and absorb light has been investigated. The derivation of the basic equation is briefly indicated, and methods for its solution described. Results are calculated for parallel light incident normal to the surface of a plane layer. The energy reflected and transmitted by the layer is found, and from this the energy absorbed is deduced. The effects of reflections at the interfaces are considered.

These results are used to explain the apparent discrepancy between the observed optical density of an unexposed photographic emulsion layer and the result calculated by using the value of the absorption coefficient of the bulk silver halide, and neglecting all effects due to scattering. Other general results predicted by the full treatment of the problem are also shown to be in agreement with experiment.

§1. Introduction

The purpose of this paper is to calculate and compare with experiment the reflection and transmission by a layer composed of particles embedded in a transparent medium, each particle scattering and absorbing light. An essential feature of the calculation is the consideration of the effect of reflection of the light between the particles, which gives rise to the characteristically turbid appearance.

As one example of the application of these results they have been used to explain certain apparently anomalous properties of unexposed photographic emulsion layers, and to give an account of their dependence on the physical characteristics of the system. For these purposes only approximate solutions are needed, so that although the problem remains complicated, some of the formidable mathematical difficulties encountered in the derivation of highly accurate results (discussed in simpler physical problems by Kourganoff (1952)) do not arise.

Details of the properties which require explanation are given in §2. In §3 the basic equations are derived and the implicit approximations indicated. Approximate solutions are derived in §4, and the choice of numerical values for the various physical parameters describing the system is discussed in §5. Results are compared with experiment in §6.

§2. The Optical Densities of Unexposed Photographic Emulsion Layers

In a layer containing many absorbing and scattering particles an incident ray of light is reflected between the particles, and since its path length is thereby

* Communication No. 1568H from the Kodak Research Laboratories.

increased, the loss of energy suffered by the ray on passing through the layer is greater than that expected if the scattering were ignored. The particles in photographic emulsions are of course silver halide crystals. The main features of interest in the optical properties of the layer are the reflected flux R and the transmitted flux T for collimated light of unit intensity incident normal to the plane surface of the layer. If these quantities are known the optical density D of the unexposed emulsion can be calculated from the definition

$$D = \log\left(\frac{\text{net incident flux}}{\text{transmitted flux}}\right) \equiv \log\left(\frac{1-R}{T}\right) \equiv \log\left(1+\frac{A}{T}\right). \quad \ldots\ldots(2.1)$$

Here A is the energy absorbed per unit area of the layer per second for unit incident intensity. It will be noticed that D does not depend on the absorption alone.

To calculate the value of D the absorption per particle must be known. Unfortunately this can only be roughly estimated, because the individual halide crystals have many different sizes and complicated geometrical forms, and are of the same order of linear dimensions as the wavelength of the incident light. If the transmission of light of intensity I_0 through a large crystal of the halide obeys the law $I_0 \exp(-kx)$, there seems to be no more satisfactory alternative than to regard the grains as spheres of uniform radius r, and to assume that the absorption per particle per second for unit incident intensity is equal to $4\pi r^3 k/3$, where k is the absorption coefficient of the halide, for all wavelengths for which $kr \ll 1$. (This condition means that only a small fraction of the energy incident on each crystal is absorbed.) This value for the absorption per grain is of the same order as that calculated by the methods of geometrical optics, which are admittedly highly artificial in view of the smallness of the spheres. However, the assumed value is probably not in error by more than a factor of 2 at the most.

If scattering is neglected D may be found very simply. If I_0 is the incident intensity, N the number of particles per unit volume, and z_1 cm the thickness of the layer, when the above value for the absorption is used, the emergent intensity I is given by

$$I = I_0 \exp(-4\pi r^3 k N z_1/3). \quad \ldots\ldots(2.2)$$

If we write v for the volume of halide per unit area of the layer, and ignore the reflected light, the optical density is

$$D_0 = 0{\cdot}4343\, kv. \quad \ldots\ldots(2.3)$$

Because of the uncertainty in the value of the absorption per grain, this result is only approximate.

Nevertheless, as Farnell (1954, this issue of *Proceedings*) and others have pointed out, this theoretical result is undoubtedly far too small. Results by Meidinger (cited by Stenger and Staude (1944), pp. 32, 33), which have been thoroughly substantiated by Farnell, show this clearly. For example, for one emulsion with $1{\cdot}3 \times 10^{-3}$ g silver bromide per unit area the reflection and transmission at a wavelength of 4358 Å were 0·46 and 0·12 respectively, giving 0·653 as the value of D. Using eqn (2.3) and values of k from measurements by Slade and Toy (1920), D_0 is found to be equal to 0·042, about one-fifteenth of the observed value. Such a result is typical for wavelengths greater than about 4200 Å. Discrepancies of this order are far greater than possible errors introduced by ignoring the reflected flux and the uncertainty in the absorption per grain.

A convenient measure of the effect requiring explanation is afforded by the 'scatter factor' ϕ_e, defined as the ratio

$$\phi_e = \frac{\text{observed density}}{\text{calculated density neglecting scattering}}. \qquad \ldots\ldots(2.4)$$

It will be shown that the experimental results can be satisfactorily explained by including the effect of scattered light. The optical properties of the unexposed layers can then be correctly described. In particular, discrepancies such as those between calculated and observed values of D_i corresponding to large values of ϕ_e, are removed. It might have been suggested that the cause of the error was the under-estimation of the absorption per particle. This hypothesis will be shown to be unnecessary.

§3. The Equation of Transfer

The equation of transfer of radiation has been derived by many authors (for example, Chandrasekhar 1950), and so it will only be necessary to state the notation to be used and other details peculiar to the present problem.

It is assumed that the layer has thickness z_1 cm and that the surface area perpendicular to the direction of propagation of the incident collimated beam is very large compared with z_1^2; also that there are N particles randomly distributed per unit volume. Since randomness is assumed, the radiation field can be sufficiently specified by its intensity alone; the hypothesis also implies that high concentrations of particles are excluded from consideration, since here the packing effect would introduce short-range regularity throughout certain regions. The supporting medium is assumed not to scatter or absorb light.

All the particles are assumed to be spheres of radius r, equal to the mean in the specimen used for experiment. The energy absorbed per second per unit incident intensity by a particle will be written $\alpha\pi r^2$, and the energy scattered out of the direction of propagation into other directions $\sigma\pi r^2$. Thus the total loss of energy for a ray in the direction of propagation is $(\alpha+\sigma)\pi r^2$. Here α and σ also depend on r. If a wave of intensity I strikes the particle, it is assumed that the intensity of scattered light is the same for any ray on the surface of a cone with the initial ray as axis, and the intensity of a ray scattered into solid angle $d\omega$ about a direction at an angle Θ to the original forward direction is written $(\sigma\pi r^2)\gamma(\Theta)\,d\omega/4\pi$ where

$$\frac{1}{4\pi}\int \gamma(\Theta)\,d\omega = \frac{1}{2}\int_0^\pi \gamma(\Theta)\sin\Theta\,d\Theta = 1, \qquad \ldots\ldots(3.1)$$

the integral being over the unit sphere. This ensures that the total energy scattered into all directions per second is $\sigma\pi r^2$. The angular function γ is essentially positive, and describes the type of scattering. It is a complicated function of particle size and shape, the wavelength of the light, and the optical constants of the particle and supporting medium. In general we may write

$$\gamma(\Theta) = \sum_0 \omega_s P_s(\cos\Theta) \qquad \ldots\ldots(3.2)$$

where P_s is a Legendre polynomial of degree s and ω_s is a constant; in particular ω_0 is equal to 1. This form for γ satisfies the relation (3.1). Particular values of the coefficients will be discussed later.

The upper illuminated surface of the layer is taken as the x, y plane containing the origin, and the positive z axis points into the layer. The incident collimated light is thus propagated in the positive z direction. Any direction at a point in the layer is completely specified if the polar coordinates θ, ϕ of the direction parallel to it in the original reference system are given. For convenience the radiation field will be divided into two parts, one representing the passage of the collimated beam through the layer, and the other that due to light which has been scattered. If $I(z)$ is the intensity of the collimated beam at depth z, the change in intensity on going to a depth $z+dz$ is given by the loss of energy in the beam due to the particles encountered in the length dz, i.e.

$$dI = -I(z)(\alpha+\sigma)\pi r^2 N\, dz, \qquad \ldots\ldots(3.3)$$

whence
$$I = I_0\, e^{-\tau} \qquad \ldots\ldots(3.4)$$

where
$$\tau = (\alpha+\sigma)\pi r^2 N z \qquad \ldots\ldots(3.5)$$

and I_0 is the incident intensity.

If we consider a certain direction at a depth of z in the layer, parallel to the direction θ, ϕ and, at a distance dl along it, find the change dS in the intensity S of the scattered light in a small solid angle about the given direction, we obtain

$$dS(\theta, \phi, z) = -S(\theta, \phi, z)(\alpha+\sigma)\pi r^2 N\, dl + I_0\, e^{-\tau}\sigma\pi r^2 N\, dl\, \gamma(\Theta)/4\pi + \frac{\sigma\pi r^2 N\, dl}{4\pi}\int S(\theta', \phi', z)\gamma(\Theta)\, d\omega', \qquad \ldots\ldots(3.6)$$

where Θ is the angle between directions θ, ϕ and θ', ϕ'. Writing

$$\cos\theta = \mu, \quad \cos\theta' = \mu', \qquad \ldots\ldots(3.7)$$

we have
$$dz = \mu\, dl, \qquad \ldots\ldots(3.8)$$

and we may write $\gamma(\mu', \phi';\ \mu, \phi)$ for $\gamma(\Theta)$. Hence

$$\mu\frac{\partial S}{\partial \tau} = -S + \frac{\epsilon I_0 e^{-\tau}}{4\pi}\gamma(\mu, \phi;\ 1, 0) + \frac{\epsilon}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi} S(\mu', \phi', \tau)\gamma(\mu', \phi';\ \mu, \phi)\, d\mu'\, d\phi' \qquad \ldots\ldots(3.9)$$

where
$$\epsilon = \sigma/(\sigma+\alpha). \qquad \ldots\ldots(3.10)$$

If there is no absorption ϵ is unity. Since the light is incident normal to the plane surface of the layer, it is obvious from symmetry that S is independent of ϕ. Using the addition theorem for spherical harmonics, we have

$$\gamma(\mu, \phi;\ \mu', \phi') = \sum_0^s \omega_s \left[P_s(\mu)P_s(\mu') + 2\sum_1^s \frac{(s-m)!}{(s+m)!} P_s^m(\mu)P_s^m(\mu')\cos m(\phi-\phi') \right], \qquad \ldots\ldots(3.11)$$

and hence

$$\mu\frac{\partial S}{\partial \tau} = -S + \frac{\epsilon I_0 e^{-\tau}}{4\pi}\sum \omega_s P_s(\mu) + \tfrac{1}{2}\epsilon\sum\omega_s P_s(\mu)\int_{-1}^{1} S(\mu', \tau)P_s(\mu')\, d\mu'. \qquad \ldots(3.12)$$

This is the equation of transfer in a form suited to the symmetry of the present problem.

If the interfaces between the air and the medium in which the particles are dispersed are regarded as non-reflecting, the boundary conditions for S are

$$\left.\begin{aligned} S(\theta, 0) &= 0, \quad (0 \leqslant \theta < \tfrac{1}{2}\pi) \\ S(\theta, \tau_1) &= 0, \quad (\tfrac{1}{2}\pi < \theta \leqslant \pi) \end{aligned}\right\} \qquad \ldots\ldots(3.13)$$

where
$$\tau_1 = (\alpha+\sigma)\pi r^2 N z_1. \qquad \ldots\ldots(3.14)$$

The first equation in (3.13) expresses the fact that no scattered light crosses the upper surface in the forward direction, and the second that no scattered light enters the layer across the lower surface.

If the interface reflects light, with a reflection coefficient ρ, then we may write

$$\left.\begin{aligned} S(\theta, 0) &= \rho S(\pi-\theta, 0), \\ \rho S(\theta, \tau_1) &= S(\pi-\theta, \tau_1), \end{aligned}\right\} (0 \leqslant \theta < \tfrac{1}{2}\pi). \qquad \ldots\ldots(3.15)$$

Strictly ρ should also be a function of the angle, but in view of the accuracy intended in the present treatment such a refinement is unnecessary. Also in this case the collimated beam will be reflected at the interfaces. To avoid undue complication this too will be neglected, so that the results will give a lower limit to the amount of energy absorbed in the layer. This point will be considered again in §6.

Corresponding to the value of the scatter factor (2.4) derived from experiment, the theoretical value ϕ of the scatter factor may be defined as

$$\phi = \frac{\text{calculated density including effect of scattering}}{\text{calculated density neglecting scattering}}.$$

If scattering is neglected, from (3.4) it follows that the emergent intensity is I, where

$$I = I_0 \exp(-\alpha\pi r^2 N z_1) = I_0 \exp[-(1-\epsilon)\tau_1]. \qquad \ldots\ldots(3.16)$$

Hence

$$\phi = D/0{\cdot}4343(1-\epsilon)\tau_1. \qquad \ldots\ldots(3.17)$$

§4. Approximate Solutions

Two methods will be described, the first corresponding to the Eddington–Milne approximation (see Kourganoff 1952, p. 87) and the second to the method used by Mark. The calculations described in §5 are in the main based on the first method, the Mark approximation being used as a means of indicating the magnitude of the errors in the simpler treatment.

4.1. *The Eddington–Milne Method*

If we define

$$F^*(\tau) = \int_0^{2\pi}\int_{-1}^{1} S(\mu, \tau)\mu \, d\mu \, d\phi, \qquad \ldots\ldots(4.1.1)$$

$$J^*(\tau) = \int_0^{2\pi}\int_{-1}^{1} S(\mu, \tau) \, d\mu \, d\phi, \qquad \ldots\ldots(4.1.2)$$

then F^* is proportional to the net flux of scattered radiation and r^2J^* is proportional to the energy falling on a grain from all directions per second. In the usual way, by integrating (3.12) over solid angle 4π and making use of the orthogonality relations of Legendre polynomials, we obtain

$$\frac{dF^*}{d\tau} = -(1-\epsilon)J^* + \epsilon I_0 \, e^{-\tau}, \qquad \ldots\ldots(4.1.3)$$

and after multiplying (3.12) by μ and again integrating,

$$\frac{d}{d\tau}\int_0^{2\pi}\int_{-1}^{1} \mu^2 S \, d\mu \, d\phi = -\left(1 - \frac{\epsilon\omega_1}{3}\right) F^* + \frac{\epsilon\omega_1}{3} I_0 \, e^{-\tau}. \qquad \ldots\ldots(4.1.4)$$

In the integrand in eqn (4.1.4) μ^2 is now replaced by its mean value over the range $(-1, 1)$, viz. $1/3$, and hence, approximately,

$$\frac{dJ^*}{d\tau} = -(3-\epsilon\omega_1)F^* + \epsilon\omega_1 I_0 \, e^{-\tau}. \qquad \ldots\ldots(4.1.5)$$

Two cases need separate consideration.

(*a*) *Absorbing particles.* In this case $\epsilon \neq 1$, and the solutions of (4.1.3) and (4.1.5) are

$$J^* = Ae^{-\kappa\tau} + Be^{\kappa\tau} + I_0 e^{-\tau}\left(\frac{\kappa^2-3}{1-\kappa^2}\right), \qquad \ldots\ldots(4.1.6)$$

$$F^* = \tfrac{1}{2}p(Ae^{-\kappa\tau} - Be^{\kappa\tau}) - I_0 e^{-\tau}\left(\frac{1+\omega_1-\epsilon\omega_1}{1-\kappa^2}\right), \qquad \ldots\ldots(4.1.7)$$

where
$$p = 2\kappa/(3-\epsilon\omega_1), \qquad \ldots\ldots(4.1.8)$$

$$\kappa^2 = (1-\epsilon)(3-\epsilon\omega_1). \qquad \ldots\ldots(4.1.9)$$

From the boundary conditions (3.13)

$$F^*(0) = 2\pi\int_{-1}^{0} S(\mu, 0)\mu\, d\mu. \qquad \ldots\ldots(4.1.10)$$

In the integrand μ is again replaced by its mean value over the range, viz. $-\frac{1}{2}$, whence, approximately,

$$F^*(0) = -\pi\int_{-1}^{0} S(\mu, 0)\, du = -\tfrac{1}{2}J^*(0). \qquad \ldots\ldots(4.1.11')$$

In a similar way

$$F^*(\tau_1) = \tfrac{1}{2}J^*(\tau_1). \qquad \ldots\ldots(4.1.11'')$$

There relations would only be exact if

$$\left.\begin{aligned} S(\mu, \tau) &= S_+(\tau), \qquad (0 < \mu \leqslant 1) \\ S(\mu, \tau) &= S_-(\tau), \qquad (-1 \leqslant \mu < 0) \end{aligned}\right\} \qquad \ldots\ldots(4.1.12)$$

where S_+ and S_- are independent of μ, in which case (4.1.5) would also be exact. Thus the approximations which have been used are mathematically equivalent to assuming the relations (4.1.12) and using the boundary conditions (3.13).

The equations to determine A and B are

$$A(1+p) + B(1-p) = I_0\epsilon\left(\frac{5+3\omega_1-3\epsilon\omega_1}{1-\kappa^2}\right), \qquad \ldots\ldots(4.1.13)$$

$$A(1-p)\exp(-\kappa\tau_1) + B(1+p)\exp(\kappa\tau_1) = I_0\epsilon\left(\frac{1-\omega_1+\epsilon\omega_1}{1-\kappa^2}\right)\exp(-\tau_1), \quad (4.1.14)$$

and the flux R reflected for unit incident intensity is given by

$$I_0 R = \frac{1}{2}\left[A + B - I_0\left(\frac{3-\kappa^2}{1-\kappa^2}\right)\right] \qquad \ldots\ldots(4.1.15)$$

and the flux T transmitted for unit incident intensity by

$$I_0 T = \frac{1}{2}\left[A\exp(-\kappa\tau_1) + B\exp(\kappa\tau_1) - I_0\exp(-\tau_1)\left(\frac{1+\kappa^2}{1-\kappa^2}\right)\right]. \qquad \ldots\ldots(4.1.16)$$

Results are computed from (4.1.13), (4.1.14), (4.1.15) and (4.1.16) by inserting numerical values for the physical parameters. Certain limiting cases are of special interest.

(i) 'Thin' layer. If τ_1 is small, it is easily shown that

$$\frac{1-R}{T} = 1 + (1-\epsilon)\tau_1 + O(\tau_1^2). \qquad \ldots\ldots(4.1.17)$$

If scattering is entirely ignored, the emergent flux is given by (3.16) which, when expanded, gives a result identical with (4.1.17). This is to be expected, since if the

layer were made very nearly transparent (small τ_1) multiple scattering of the light could have little effect on the value of D.

(ii) ' Thick ' layer. If τ_1 and $\kappa\tau_1$ are both large, to a good approximation we have

$$\frac{B}{A} = -\left(\frac{1-P}{1+P}\right)\exp(-2\kappa\tau_1). \qquad \ldots\ldots(4.1.18)$$

From (4.1.6) it is easily shown that J^* decays practically exponentially with increasing depth throughout most of the layer. Hence the energy received by a particle, i.e. its exposure in the photographic case, decreases exponentially. This agrees with experimental observations by Farnell (private communication) on thick layers.

It follows that

$$\frac{1-R}{T} = \frac{\exp(\kappa\tau_1).(1+p)(5+3\kappa)(1-\kappa)}{2\epsilon(5+3\omega_1-3\epsilon\omega_1)}, \qquad \ldots\ldots(4.1.19)$$

and hence

$$\phi = \left(\frac{3-\epsilon\omega_1}{1-\epsilon}\right)^{1/2} + O(\tau_1^{-1}). \qquad \ldots\ldots(4.1.20)$$

Since α is proportional to the bulk halide absorption coefficient k, and in many cases of practical interest $\alpha \ll \sigma$ so that τ_1 is very little altered by changes in α, under these conditions we may write approximately

$$\phi k^{1/2} = \text{const.} \qquad \ldots\ldots(4.1.21)$$

Hence, the smaller k the greater the scatter factor in these circumstances. Using less exact methods, Mr. P. C. Burton (private communication) has suggested a result similar to this.

(iii) "Thick ' layer, small absorption. In this case κ will be small.

If $\kappa\tau_1$ is large, then, approximately,

$$(1-R)/T = \tfrac{1}{2}\exp(\kappa\tau_1). \qquad \ldots\ldots(4.1.22)$$

Case (iii) is physically very similar to the transmission by the layer of a completely diffuse incident flux. This can be calculated by putting I_0 equal to zero in (4.1.6) and (4.1.7) and writing for the incident intensity the constant value I^*/π for all angles between 0 and $\pi/2$. The incident flux is then I^*, and from the definitions (4.1.1) and (4.1.2), and employing approximations similar to those in (4.1.10), the boundary conditions in this case become

$$\left.\begin{aligned} F^*(0) + \tfrac{1}{2}J^*(0) &= 2I^*, \\ F^*(\tau_1) - \tfrac{1}{2}J^*(\tau_1) &= 0. \end{aligned}\right\} \qquad \ldots\ldots(4.1.23)$$

Solving (4.1.6), (4.1.7) and (4.1.22), having taken I_0 to be zero, we find

$$\frac{1-R}{T} = \frac{F^*(0)}{F^*(\tau_1)} = \tfrac{1}{2}[(1+p)\exp(\kappa\tau_1) + (1-p)\exp(-\kappa\tau_1)]. \qquad \ldots\ldots(4.1.24)$$

Hence when $\kappa\tau_1$ is large, and κ is small, approximately

$$(1-R)/T = \tfrac{1}{2}\exp(\kappa\tau_1), \qquad \ldots\ldots(4.1.25)$$

in agreement with (4.1.22).

When the boundary conditions (3.15) are used, the case of reflections at the interface can be examined in a similar way to the above. Some care is necessary in calculating the reflected and transmitted flux, but no essentially new difficulties arise. As already stated, the reflection by the lower interface of the collimated beam back through the layer towards the upper interface is neglected. Details

of the work are straightforward but tedious and have been omitted for brevity. The results have been used for numerical work which is discussed in § 6.

(*b*) *Non-absorbing particles.* In this case $\epsilon = 1$ and the solution of (4.1.3) and (4.1.5.) is

$$J^* = -3I_0 e^{-\tau} - (3-\omega_1) L\tau + M, \qquad \ldots\ldots(4.1.26)$$

$$F^* = -I_0 e^{-\tau} + L. \qquad \ldots\ldots(4.1.27)$$

From (4.1.27) it will be seen that the net flux due to both collimated light and scattered light is simply L, that is, a constant independent of τ. This result is of course well known. Using (4.1.10) and (4.1.11) we find the reflected and total transmitted flux for unit incident intensity to be

$$R = \frac{\exp(-\tau_1) + (3-\omega_1)\tau_1 - 1}{4 + (3-\omega_1)\tau_1}, \qquad \ldots\ldots(4.1.28)$$

$$T = \frac{5 - \exp(-\tau_1)}{4 + (3-\omega_1)\tau_1}. \qquad \ldots\ldots(4.1.29)$$

The case of reflection at the interfaces can be treated in a similar way.

4.2. *Mark's Method*

This method (which is described in a ' declassified report ', see bibliography) is very tedious when many examples are to be considered, and each requires separate computation. However, it does provide a method of increasing the accuracy of the solution. The intensity of the scattered light is written

$$S = \tfrac{1}{2}\sum_0^M (2n+1)\psi_n(\tau) P_n(\mu), \qquad \ldots\ldots(4.2.1)$$

where ψ_n is a function of τ alone. On multiplication of (3.12) by $P_m(\mu)$ $(0 \leqslant m \leqslant M)$ and integration with respect to μ between -1 and $+1$, using (4.2.1) and the recurrence and orthogonality relations of Legendre polynomials, there results

$$(m+1)\psi'_{m+1} + m\psi'_{m-1} = -f_m\psi_m + \frac{\epsilon I_0 \omega_m}{2\pi} e^{-\tau} \qquad \ldots\ldots(4.2.2)$$

where

$$f_m = 2m + 1 - \epsilon\omega_m \qquad \ldots\ldots(4.2.3)$$

and dashes denote differentiation with respect to τ. It can easily be shown that f_m is never negative (see Appendix). As an approximation, in the equation of the set (4.2.2) with $m = M$, the term involving ψ'_{M+1} is dropped, so that the last equation of the set is

$$M\psi'_{M-1} = -f_M\psi_M + \frac{\epsilon I_0 \omega_M}{2\pi} e^{-\tau}. \qquad \ldots\ldots(4.2.2')$$

It is then possible to obtain ψ_n by the usual means, and in fact

$$\psi_n = \sum A_s \alpha_{s,n} \exp(\lambda_s \tau) + c_n \exp(-\tau). \qquad \ldots\ldots(4.2.4)$$

Here c_n, $\alpha_{s,n}$ and λ_s are (determined) constants, and the A_s are to be found from the boundary conditions. The summation is over all possible values of s, the λ_s being roots of the determinant

$$\begin{vmatrix} f_0 & \lambda & & & & \\ \lambda & f_1 & 2\lambda & & 0 & \\ & 2\lambda & f_2 & 3\lambda & & \\ & & \cdot & \cdot & & \\ & & \cdot & \cdot & & \\ & 0 & & (M-1)\lambda & f_{M-1} & M\lambda \\ & & & & M\lambda & f_M \end{vmatrix} = 0. \qquad \ldots\ldots(4.2.5)$$

It is easily shown that the roots occur in pairs (positive and negative) and are $M+1$ (M) in number, if M is odd (even). Thus there are $M+1$ or M constants $A_{s,n}$ depending on whether M is odd or even.

As in other problems to which this method has been applied, the unsatisfactory nature of the form (4.2.1) assumed for S becomes clear when the boundary conditions are considered. Using (3.13) we must have for all n

$$\left.\begin{aligned}\int_0^1 P_n(\mu)\sum_0^M(2l+1)\psi_l(0)P_l(\mu)\,d\mu &= 0\\ \int_{-1}^0 P_n(\mu)\sum_0^M(2l+1)\psi_l(\tau_1)P_l(\mu)\,d\mu &= 0\end{aligned}\right\} \quad \ldots\ldots(4.2.6)$$

so that there are in all $2(M+1)$ conditions to be satisfied by the constants A_s. This is impossible, so the conditions (4.2.6) must be relaxed in some way. The usual way is to retain (4.2.6) for odd values of n only. Consideration of the continuity of the solutions (see work by Mark) makes this choice appear plausible. It should be noticed that with $n=1$ the equations (4.2.6) imply that the flux due to scattered light is zero in the forward direction over the upper interface, and zero in the backward direction over the lower interface.

The reflected flux R and the transmitted flux T for unit incident intensity are given by

$$\left.\begin{aligned}I_0R &= -2\pi\int_{-1}^0 P_1(\mu)S(\mu,0)\,d\mu,\\ I_0T &= 2\pi\int_0^1 P_1(\mu)S(\mu,\tau_1)\,d\mu + I_0\exp(-\tau_1).\end{aligned}\right\} \quad \ldots\ldots(4.2.7)$$

Results have been calculated from the foregoing equations, in which suitable values of the parameters have been inserted. When many different examples are to be considered the numerical work becomes very heavy.

It can easily be shown that if M is taken equal to unity, results calculated by this method are identical with those from the Eddington–Milne approximation.

In a similar way the case of non-absorbing particles ($\epsilon=1$) can be calculated; the only important difference from the above is the appearance of a term proportional to τ, just as in the simpler method. No extra difficulties arise, and details have been omitted for brevity.

§5. Numerical Values

An unexposed photographic emulsion layer will be chosen as an example of a material containing scattering and absorbing particles, which in this case are silver halide crystals. The mean value of r is found by measurement of the sample whose properties are to be compared with theory; a very rough average order of magnitude for r is about $0{\cdot}25\times10^{-4}$ cm for an emulsion of pure silver bromide crystals. If the mean value of r and the weight of halide per unit area of the layer are known, an effective value of the product Nz_1r^2 occurring in the definition of τ_1 can be found with moderate accuracy. However, by the nature of the system, values of α, σ and γ are much less certain. These physical parameters describe essential properties of the individual particles, and must be known before numerical work is possible.

Because of the non-uniformity and smallness of the particles it is difficult to find any convincing way of improving the accuracy of the value $4\pi r^3k/3$, assumed for the mean absorption per second by a particle for unit incident

intensity. With this assumption $\alpha = 4rk/3$. The value of α is strongly dependent on the wavelength, but for wavelengths greater than about 4000 Å, $\alpha < 0{\cdot}05$ (see table 3). For perfect spheres values of σ are known from the work of Mie (1908; numerical values tabulated by Lowan 1949) and for the size of particles quoted above, with optical constants corresponding to those for silver halide embedded in gelatine, and incident radiation in the visible region, σ is between 2 and 3. Although for perfect spheres σ depends on the wavelength of the incident light, it is probable that for our purposes it should be regarded as effectively independent of wavelength, since the particles in the practical example are non-uniform, and the value of σ for each size-class would be changing with wavelength in a different way. It would be difficult to suggest any simple alternative which is demonstrably more accurate. Just as in the estimation of α, it is again the non-uniformity of the particles which causes the unavoidable uncertainty. It is clear, however, that $\sigma \gg \alpha$, and hence changes in α due to changes in wavelength in the visible region leave $\sigma + \alpha$ practically unchanged. The changes in α are of course very important, since they determine the value of κ^2, which by (4.1.9) is seen to be proportional to $\alpha/(\sigma + \alpha)$.

The constants ω_s in the function γ must be stated. These also depend on the particle size and shape, and so again only very crude average values can be given. If the incident light travels in the direction $\theta = 0$, the main feature of the function for particles of the size in question is that the intensity of scattered light is greatest for small angles, and decreases very rapidly with increasing θ, becoming almost negligible for angles greater than $\pi/2$, i.e. scattering is predominantly in forward directions. By a rough estimation of the values of the constants appropriate to an average particle the corresponding intensity distribution could be found in tables of scattering functions (Lowan 1949) and the constants ω_s calculated in the usual way. The result was

$$\gamma(\mu) = 1 + 1{\cdot}89P_1(\mu) + 1{\cdot}30P_2(\mu) + 0{\cdot}48P_3(\mu) + 0{\cdot}13P_4(\mu) + 0{\cdot}02P_5(\mu). \quad \ldots(5.1)$$

Calculations were also made by the Eddington–Milne method, assuming that ω_1 is equal to unity. In this way some indication was obtained of the extent to which the calculated results depend on the value of ω_1.

When reflections at the interfaces are included, the value 0·5 was taken for ρ. This apparently rather large value is to compensate for the total internal reflection of the light from the interior of the layer, which occurs for all angles of incidence greater than the critical value.

Knowledge of α, σ, γ and ρ is of course basic for any calculation of the optical properties of the system. From the foregoing discussion it will be obvious that in any practical example these quantities can never be stated with great accuracy. This means that it is always difficult to say precisely to which theoretical model any practical case corresponds. Rather than attempt a comparison of an experimental result with a theoretical prediction, it is better to calculate many different examples, and so to obtain a general picture of the expected dependence on the several parameters. The position of the experimental result in this scheme can then be estimated.

§6. Comparison with Experiment

By relatively simple integration, which need not be reproduced, it may be shown that the energy falling on a particle from all directions, arising from light which has been scattered once, is proportional to τ_1. The energy from light which has been

scattered twice is proportional to τ_1^2. This gives a rough means for deciding when multiple scattering is likely to be of importance : when $\tau_1 \ll 1$ scattering may be neglected; if $\tau_1 \geqslant 1$, its effect must be considered.

First, it is of interest to examine the results in the case of non-absorbing particles ($\epsilon = 1$). Reflections at the interfaces are neglected, and the value of γ given by (5.1) is assumed. The transmitted flux T is given in table 1 for various

Table 1. Transmission by a Layer of Non-Absorbing Particles

τ_1	1	2	4	8	16
T_1	0·9065	0·7821	0·5902	0·3882	0·2298
T_3	0·8895	0·7655	0·5806	0·3853	0·2298
T'	0·8895	0·8010	0·6681	0·5016	0·3345

values of τ_1, the subscripts 1 and 3 referring to results obtained by the Mark method in which M is taken equal to these values respectively. The value T_1 is also found if the Eddington–Milne method is used. From the table it will be seen that the difference between the approximations is relatively small.

An alternative way of finding the properties of thick layers of non-absorbing particles is to regard the thinnest layer, of known properties, as a unit from which the thick layers may be built by placing many such layers in optical contact. Writing t_n for the transmission of a pile of n such layers, by considering inter-reflections in the elementary way, it is easily shown that

$$\frac{1}{t_{n+1}} - \frac{1}{t_n} = \frac{1-t_1}{t_1}, \qquad \ldots\ldots(6.1)$$

so that

$$T' = t_n = \frac{t_1}{n-(n-1)t_1}. \qquad \ldots\ldots(6.2)$$

Taking t_1 equal to the value 0·8895 as shown in table 1, the transmission T' can be calculated for the different values of n. It will be seen that the error is large and steadily increases. The value of t_1 required if T' is to agree with the tabulated value when n is equal to 16 is 0·827, which differs from the tabulated result by an amount greater than possible errors due to approximations. The simple method of treating the thick layer as a pile of thinner ones therefore gives unreliable results, with a large error when the number of thin layers is large, because it ignores changes in the dependence of intensity on direction.

For absorbing particles the method based on the Eddington–Milne approximation has been used to calculate the reflection and transmission for many different values of τ_1 and ϵ. Two functions have been chosen for γ, with ω_1 equal to 1·89 (see 5.1) and 1·0 respectively. From these results D has been calculated and the scatter factor ϕ defined by (3.17) derived. Figures 1 and 2 show these results; ϕ has been plotted as a function of D, and the effect of variation of τ_1 and ϵ is clearly shown by the course of the curves. The asymptotic value of ϕ given by (4.1.20) for large τ_1 has also been inserted. The figures show that particularly large values of ϕ occur when $1-\epsilon$ is small and τ_1 large. The latter condition has already been shown to imply that scattering is important ; the former means that $\alpha \ll \sigma$, i.e. that most of the light falling on a particle is scattered, so that in its subsequent passage through the layer it will fall on other particles, and again suffer loss of energy by absorption. It is also of interest to examine the quantity f, where

$$f = \frac{1-R-T}{1-\exp[-(1-\epsilon)\tau_1]}, \qquad \ldots\ldots(6.3)$$

that is, the ratio of energy absorbed to the calculated absorption, assuming that scattering is negligible. The quantity has been plotted in figure 3 as a function of D, from which it is clear that in conditions giving a large scatter factor ϕ the energy absorbed is greater than that predicted by the elementary treatment ignoring scattering. However, the increase is seldom more than a factor of 2, in contrast to the factor ϕ, which is often much greater. With increasing thickness f tends to the value $1-R$.

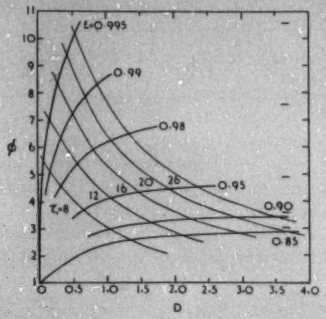


Figure 1. ϕ as a function of D, showing the effect of change of τ_1 and ϵ; $\omega_1=1{\cdot}89$. Asymptotic values of ϕ when $\tau_1=\infty$ are shown at extreme right.

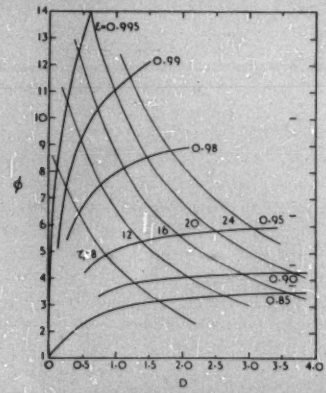


Figure 2. ϕ as a function of D, showing the effect of change of τ_1 and ϵ; $\omega_1=1{\cdot}0$. Asymptotic values of ϕ when $\tau_1=\infty$ are shown at extreme right.

In table 2 the effects of reflections at the interfaces are shown. The value 0·5 has been chosen for ρ, and the calculation made for the case in which ω_1 is equal to unity. Comparison with figure 2 shows that a considerable increase in ϕ occurs for a given density compared with the value when ρ is zero. In these calculations the fact that the collimated incident beam is also reflected backwards through the layer by the lower interface has been neglected. If this were included, the

values of ϕ would be still more increased. Similar results are found for the calculation with ω_1 equal to 1·89. These results are in complete agreement with the qualitative ideas about the effect of scattering discussed above.

The theoretical results have therefore demonstrated the possibility of large values of ϕ under certain conditions. It remains to show that these conditions in fact correspond roughly to those in which large scatter factors are observed experimentally.

As already explained in § 5, the form of the function γ is uncertain. It seems likely that the value to be taken for ω_1 is less than that suggested by the Mie theory for spheres of size equal to the mean, since in practice many smaller spheres will be present, for which the scattering is less strongly in forward directions, and the larger particles will be of irregular shape. Hence the average scattering pattern can be expected to show less predominance in forward directions, which means that

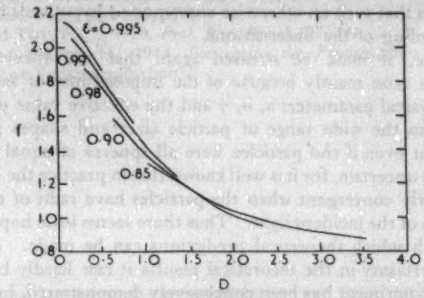


Figure 3. f as a function of D, showing the effect of change of ϵ; $\omega_1 = 1{\cdot}89$.

Table 2. Values of D and ϕ when $\epsilon = 0{\cdot}995$ and Reflections at the Interfaces are included

τ_1	8	12	16	20	24
D	0·195	0·343	0·506	0·674	0·846
ϕ	11·2	13·2	14·6	15·5	16·2

ω_1 is reduced (see Appendix). Probably the value unity is a lower limit. Values of τ_1 are also uncertain and, moreover, vary considerably from case to case but, on the assumption of a value of 2 for $(\alpha+\sigma)$, they usually lie between about 12 and 20, which is a range of values for which results are shown in the figures. In the particular example quoted in § 2 τ_1 is equal to 13, which is somewhat lower than the average.

Taking a value of $0{\cdot}2 \times 10^{-4}$ cm for the mean radius, as in the example of § 2, table 3 shows the variation of $\alpha = 4rk/3$ with wavelength for crystals of pure silver bromide. For a wavelength of 4358 Å $\alpha/(\sigma+\alpha)$ is thus about 0·006, and hence ϵ is about 0·994. Because of all the uncertainties previously described, the difference between this value and 0·995 is not significant, and we may compare results calculated for the latter value with the experimental observations.

When τ_1 is about 13, and ϵ about 0·995, from figure 1 ϕ is approximately 7·5. From figure 2 the value is 11, and from table 2 roughly 13·5. Experimentally ϕ_0 was equal to 15, a value itself subject to not inconsiderable errors owing to

practical difficulties. Thus it appears that when all effects are included the calculated value of ϕ is not greatly different from the observed value. Similar general results are found in all the other examples examined, the calculated values usually appearing somewhat too small, as in the above case. An explanation of the large values of ϕ has thus been found in the effects of scattering.

Table 3. Dependence of α on Wavelength, for Pure Silver Bromide

λ(Å)	4047	4100	4200	4300	4358	4400	4500	4600
$10^3\alpha$	4·3	3·7	2·4	1·6	1·3	1·1	0·66	0·25

It could have been argued that these large values arose because the value of k to be taken in (2.3) had been under-estimated, that is, the particles were really absorbing many times more strongly than suggested by simple considerations. A fifteen-fold increase in absorption would be necessary to remove the discrepancy in the example quoted in §2. However, it is clear from the theoretical results that such an otherwise unsupported hypothesis is not necessary for an understanding of the observations.

In conclusion, it must be stressed again that the uncertainties in the calculated results arise mainly because of the impossibility of assigning precise values to the physical parameters α, σ, γ and the effective value of Nz_1r^2. This is partly owing to the wide range of particle sizes and shapes present in any practical case, but even if the particles were all spheres of equal size, the value of α would still be uncertain, for it is well known that in practice the Mie expression for α is very slowly convergent when the particles have radii of the same order as the wavelength of the incident light. Thus there seems little hope of improving the precision with which theoretical predictions can be made. In view of the unavoidable uncertainty in the theoretical results it can hardly be claimed that agreement with experiment has been conclusively demonstrated, but it does seem that there is no definite incompatibility. In particular the surprisingly large values of ϕ are fully explained by the consideration of the effect of scattering of the light.

ACKNOWLEDGMENTS

The author wishes to thank Mr. P. C. Burton, Dr. G. C. Farnell and Mr. E. W. H. Selwyn, all of the Research Laboratories, Kodak Ltd., for their suggestion of a theoretical investigation of this problem, and for many helpful and stimulating discussions.

APPENDIX

Writing μ for $\cos\theta$, from (3.2) we have

$$\gamma(\mu)=\sum_0 \omega_s P_s(\mu). \qquad \text{.....(A 1)}$$

Hence

$$\int_{-1}^{1}\gamma(\mu)P_m(\mu)\,d\mu=\frac{2\omega_m}{2m+1}. \qquad \text{.....(A 2)}$$

Since $\gamma(\mu)$ is everywhere positive or zero and $P_m(\mu)$ is less than or equal to 1, it follows that

$$\frac{2\omega_m}{2m+1}\leqslant\int_{-1}^{1}\gamma(\mu)\,d\mu=2. \qquad \text{.....(A 3)}$$

Thus

$$2m+1-\epsilon\omega_m=f_m\geqslant 0, \qquad (\epsilon\leqslant 1). \qquad \text{.....(A 4)}$$

It should also be noticed that from (A 2) the value of ω_1 is given by

$$\frac{2\omega_1}{3} = \int_{-1}^{1} \gamma(\mu)\mu \, d\mu = \int_{0}^{1} [\gamma(\mu) - \gamma(-\mu)]\mu \, d\mu. \qquad \ldots\ldots(A\,5)$$

Since $\gamma(\mu)(0 \leqslant \mu \leqslant 1)$ is the intensity in a forward direction, and $\gamma(-\mu)$ is the intensity in a backward direction, if scattering is increasingly in forward directions, the expression $[\gamma(\mu) - \gamma(-\mu)]$ increases, and thus ω_1 increases.

References

Chandrasekhar, S., 1950, *Radiative Transfer* (Oxford : University Press).

Farnell, G. C., 1954, *Proc. Phys. Soc.* B, **67**, 164.

Kourganoff, V., 1952, *Basic Methods in Transfer Problems* (Oxford : University Press).

Lowan, A. N., 1949, *Tables of Scattering Functions for Spherical Particles* (Department of Commerce, U.S. National Bureau of Standards), *Applied Mathematics Series* 4.

Mark, C., *Declassified Montreal Reports*, C.R.T. 338, 340.

Mie, G., 1908, *Ann. Phys., Lpz.*, **25**, 377.

Slade, R. E., and Toy, F. C., 1920, *Proc. Roy. Soc.* A, **97**, 181.

Stenger, E., and Staude, H., 1944, *Fortschritte der Photographie*, Vol. 3 (Leipzig : Akademische Verlagsgesellschaft).

The Dynamic Mechanical Properties of some Organic Glasses

By J. J. BENBOW

Imperial Chemical Industries Limited, Butterwick Research Laboratories, Welwyn, Herts.

MS. received 9th October 1953

Abstract. The dynamic shear properties of two organic glasses, 2′-hydroxy-2:4:4:6:5′-pentamethylflavan and glycerol sextol (cyclo-hexanol) phthalate, have been studied at frequencies between 10^{-4} c/s and 10^{4} c/s. A brief outline is given of similar measurements on other types of materials and the measured properties of the two glasses are contrasted with those of polymethyl methacrylate in detail. Calculations have been made of the approximate relaxation-time spectra and it is shown that the behaviour of hydroxy-pentamethylflavan is governed by a single activation process over the major part of the frequency range considered. The results suggest that a correlation exists between the mechanical and chemical properties of the glasses.

§ 1. Introduction

The purpose of this paper is to describe measurements which have been made of the dynamic shear properties of two organic glasses, and to show how the results suggest a correlation between the mechanical and chemical properties of these materials. Their behaviour will be contrasted in some detail with that of polymethyl methacrylate, since this material is a glass-like high polymer on which a large number of measurements are available.

The glasses studied were 2′-hydroxy-2:4:4:6:5′-pentamethylflavan (Baker, Curtis and McOmie 1952) and glycerol sextol phthalate. The term 'organic glass' will be used here to denote liquids which do not crystallize when the temperature is reduced but steadily increase in viscosity, whatever the rate of cooling. Thus organic glasses may be glass-like solids or freely flowing liquids, depending upon their temperature. Materials which are included within this definition are not necessarily colourless, but the two chosen were, in fact, transparent and the relaxation of shear stresses in them has been studied by observing the decay of birefringence (Crawford 1953). A common organic glass is glycerol, having a viscosity of about 5 poises at 21°c which has been estimated to increase to $3{\cdot}3 \times 10^{4}$ poises at -46°c (Fox and Litovitz 1951).

2′-hydroxy-2:4:4:6:5′-pentamethylflavan is a freely flowing liquid at about 60°c but at 15°c it is a brittle solid. The chemical nature of this organic glass is quite well defined and it is unlikely that there is very much association between its molecules. This follows from the fact that it contains only one free hydroxyl group per molecule, and this is in such a position that it preferentially enters into intra-molecular hydrogen bonding rather than the inter-molecular type. Glycerol sextol phthalate is the product of esterification of phthalic acid with a mixture of glycerol and cyclo-hexanol ('sextol'). Chemically it is a rather

ill-defined substance because, depending on the conditions of preparation and the proportion of the reactants used, the degree of esterification and the amount of cross-linking can vary widely, with the result that different batches of the material rarely show reproducible physical properties. Consequently any physical measurements carried out on this glass refer only to a particular sample. However, as this glass was readily available in large quantities, it was chosen to explore this class of materials and to test the various apparatuses required. Furthermore, since it contained hydroxyl groups free to form hydrogen bonds, it provided an example of an organic glass capable of considerable intermolecular association.

The present extensive use which is being made of high polymers and similar plastic materials has provided the incentive for many workers to study the dynamic mechanical properties of these materials, for example Ferry (1953), Hillier (1949), Lethersich (1950), Mason (1951), Nolle (1950), and Nolle and Mifsud (1953). This has resulted in detailed knowledge of the mechanical behaviour of many common polymers over wide temperature and frequency ranges, and it has been shown that the response of high polymers to stresses can only be represented by an infinite number of elastic and viscous mechanisms. The relaxation-time spectra which have been calculated for high polymers are very flat and extend over several decades of frequency. This is not surprising since, on deforming any high polymer, very many different molecular mechanisms are involved. The theoretical interpretation of the experimental results on a molecular scale is consequently extremely complicated. At the same time, since the relaxation-time spectra of these materials are so broad, measurements must be made over very wide frequency ranges to show any definite trends in their behaviour.

A considerable amount of research has been carried out on the dynamic properties of inorganic glasses (e.g. Fitzgerald 1951, Jones 1948–9, Kamel 1949, Marx and Sivertsen 1953, and Weyl 1948). These experiments, made at temperatures which were well below the softening point of the glasses, showed that the relaxation times spread over a wide range and also that the spectrum had at least one maximum at low audio frequencies. At these temperatures the relaxation of elastic stresses due to viscous flow in inorganic glasses is too slow to be conveniently determined by oscillatory methods.

Two possible reasons for the observed wide distribution of relaxation times have been discussed by Jones (1948–9); they are the presence of chemical inhomogeneities, and a distribution of potential trough depths in the structure of the glasses. In many respects the spectra of inorganic glasses, at room temperature, closely resemble those of metals; Kamel's results, for example, have shown how similar the dependence of the logarithmic decrement upon frequency is, for several metals and inorganic glasses. Zener (1948) has fully discussed the interpretation of the results obtained for metals and has shown conclusively that the various peaks in the spectrum can be accounted for in terms of such effects as thermal diffusion, viscous slip at crystal boundaries and at very high frequencies when the grain size becomes comparable with the wavelength of the stress waves, thermal conduction at the crystal boundaries.

The results which have been mentioned imply that polymers and inorganic glasses, at temperatures well below their softening point, exhibit such complicated mechanical behaviour that the relation between their chemical and physical

properties tends to be obscured. This suggests that the study of compounds having as simple a chemical nature as possible should prove more informative. Such materials are the simpler unassociated hydrocarbons, carbon tetrachloride, carbon disulphide and benzene being examples. However, since at laboratory temperatures these liquids have relaxation times as short as 10^{-9} second, measurements of their dynamic shear properties would have to be made at ultra- or hyper-sonic frequencies. On the other hand, organic glasses have viscosities which increase enormously compared with the increase in the elasticities as their temperatures are reduced. Over a 30°c temperature range their relaxation times may change by as much as a millionfold. Thus the relaxation time of hydroxy-pentamethylflavan varies from 16 seconds at 6°c to 2×10^{-4} second at 30°c (Crawford 1953). Hence at sufficiently low temperatures the dynamic properties can be conveniently measured from frequencies of 10^{-4} c/s upwards.

§ 2. Experimental

The temperature range covered with hydroxy-pentamethylflavan was from about 8°c to 18°c, whilst with glycerol sextol phthalate it was from 22°c to 32°c. The frequency range within which the shear elasticity and viscosity have been measured was from 10^{-4} c/s to 7×10^{3} c/s. To cover this frequency range three different apparatuses were employed, each working over a part of the range. Preliminary measurements showed that the shear elasticity of both glasses was of the order of 10^{10} dyn cm^{-2} at the temperatures employed. This high value affected the design of the apparatus, since care had to be taken to ensure that the elasticity of any metal parts of the apparatus did not give misleading results. Because of this, the specimens were used in the form of rods which were sheared by applying torsional stresses to their ends, this method being found preferable to the more usual one of shearing between coaxial cylinders. Fortunately the two organic glasses measured adhered strongly to brass, and this property was used to bond the rods to metal end-pieces by placing the glass in contact with heated metal and allowing them to cool together. By making the diameter of these end-pieces large compared with that of the glass rod the elasticity of the metal could be safely neglected.

At low frequencies (periods 0·85 second to 21 500 seconds) the stress and strain were measured in a cylindrical rod of the organic glass. To do this a sinusoidal torsional stress was applied to the rod by means of an oscillating torsion head. The torsion head was driven by a synchronous motor through a twelve-speed gear box and a harmonic transformer which converts the motion to an oscillating rotation. The period of the oscillations could be varied in steps having ratios of $2\frac{1}{2}:1$ by selecting the appropriate gear chain. The torsion head was coupled to the top of the sample, as shown in figure 1, the lower end of the sample being bonded to a brass rod which was itself rigidly bolted to the frame of the apparatus. The temperature of the sample was maintained constant to 0·2°c by circulating a thermostatically controlled mixture of ethylene-glycol and water through the glass jacket surrounding the specimen. The small temperature difference between the sample and the controlled bath was measured by calibrated thermojunctions. To measure the strain in the glass rod a small galvanometer mirror, cemented to the upper end of the sample, was used with a conventional galvanometer lamp and scale. A second mirror at the top of the torsion head reflected light from another lamp on to the same scale, thus enabling

the motion of the torsion head to be determined as a function of time. The amplitude of the applied stress and the phase difference between the stress and strain were calculated from the motion of the two spots of light. Since the stress is proportional to the amount the torsion wire is twisted, it was necessary to subtract vectorially the motion at the top of the sample from that at the top of the torsion wire in order to obtain the true magnitude and phase of the stress.

This method is similar to that used by Lethersich in his work on solid dielectrics at roughly comparable frequencies. Recently a similar method has been described which operates at audio frequencies (Markovitz, Yavorsky, Harper, Zapas and DeWitt 1952). However, use of this instrument is complicated by the need to take the inertia of the sample into account (Markovitz 1952).

At low audio frequencies (20 c/s to 500 c/s) the dynamic shear properties were measured by observing the free decay of torsional oscillations of a mechanical resonant system. This was done by using the sample as the elastic and viscous elements of the resonant system with an added inertia which was very large

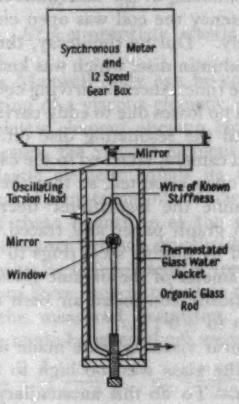


Figure 1. Low frequency apparatus.

compared with the inertia of the sample. The glasses were moulded into cylinders, 2·00 cm in diameter, 0·10–2·0 cm long, and bonded to two metal discs. The upper of these discs was screwed to the duralumin inertia disc and both were suspended from the top of the instrument by a light phosphor bronze strip (0·01 in. × 0·001 in.). The suspension wire was sufficiently strong to support the weight of the inertia disc and also sufficiently flexible to be neglected in comparison with the stiffness of the glass cylinders. The lower disc, to which the glass was bonded, was bolted rigidly to the very heavy base of the apparatus. Torsional oscillations were generated by an electrodynamic transducer consisting of a coil, mounted on a short arm attached to the inertia disc, which moved in the cylindrical gap of a permanent magnet of the type used in loudspeakers, bolted to the base of the apparatus. The coil was supplied with an alternating current of variable frequency from a low power amplifier driven by a Dawe oscillator type 400B. The torsional resonant frequency was found by varying the frequency of the oscillator until the amplitude of the mechanical oscillations was a maximum.

To observe the oscillations a condenser microphone was used since it was both highly sensitive and also did not load, appreciably, the oscillating disc. The microphone consisted of two parallel plates having an area of about one square inch. One of these was fixed to the inertia disc, with its plane radial to the disc, and the other was mounted on a micrometer thread so that the planes of both discs were accurately parallel and their distance apart could be adjusted. The micrometer thread was carried by, but insulated from, a heavy brass pillar bolted to the base of the apparatus.

As the disc vibrated the capacity of the condenser varied, and the variation was used to frequency modulate a 15 Mc/s oscillator. After being amplified the output from the oscillator was detected by a tuned diode rectifier. The 15 Mc/s component was removed from the detector circuit by a filter and, after further amplification, the audio frequency signal was displayed on one trace of a double-beam cathode-ray oscilloscope. A standard timing signal was displayed on the second beam of the oscilloscope to enable the frequency of the oscillation in the glass cylinders to be determined. With the system being driven at its resonant frequency the coil was open circuited and the oscillations were allowed to decay freely. During the decay, the frequency depended only upon the inertia of the duralumin disc, which was known, and the stiffness of the glass cylinder. At the same time, since the driving coil was open circuited during the decay period, there were no losses due to eddy currents in it or due to coupling between the transducer and the resonating disc (cf. Davies and James 1934). The shutter of a 35 mm film camera, attached to the cathode-ray oscilloscope, was opened when the coil was open circuited, so that the decaying oscillations were photographed. To determine the logarithmic decrement and frequency the films were projected on to graph paper and traced. To cover the frequency range the inertia was varied by bolting steel rings to the duralumin disc and by changing the length and diameter of the organic glass cylinder. The apparatus was placed in a thermostatically controlled air bath so that the temperature of the glass remained constant to $\pm 0{\cdot}1°$C.

Measurements of the shear viscosity were made under conditions where the logarithmic decrement of the glass was too high to form an oscillating system with only an added inertia. To do this an auxiliary elastic element of known stiffness was introduced between the inertia disc and the organic glass so that the resonant system consisted of the inertia disc, a phosphor bronze rod of negligible inertia, and the glass cylinder. The compliance of the phosphor bronze rod was chosen to be more than ten times that of the glass rod, so that the logarithmic decrement of the oscillations was always less than a tenth of what it would have been without the additional rod. First, the frequency and logarithmic decrement were found without the glass rod and with the lower end of the phosphor bronze rod rigidly clamped. Then the glass rod was placed in position and the frequency and logarithmic decrement were found as functions of temperature. The change in the frequency was very small and hence this method was not suitable for measuring the elasticity of the glasses. To alter the frequency the stiffness of this phosphor bronze rod and the inertia of the disc could be changed.

Measurements at high audio frequencies were carried out by a method using a rod of organic glass oscillating at its torsional resonant frequency. This method has been described in detail elsewhere (Benbow 1953) so that only a brief description will be given here. The glass rod forms the suspension of a ballistic moving-coil galvanometer, the coil being driven by an alternating current at a frequency

equal to the natural torsional frequency of the glass rod. The coil has insufficient inertia and stiffness to affect the frequency or damping of the oscillations. A parallel beam of light, which fell upon a photoelectric cell after being reflected from a small mirror attached to the specimen, was used to detect the oscillations. The logarithmic decrement and frequency were measured during the free decay of the oscillations; this was facilitated by repeatedly displaying the exponentially damped sine wave on a cathode-ray oscilloscope.

§ 3. Interpretation of Results

Before the measured quantities such as logarithmic decrement, phase lag and resonant frequency can be converted into values of elasticity and viscosity it is necessary to make some assumptions regarding the mechanical behaviour of the glasses. The superficial physical nature of glasses suggests that they may be expected to behave as a Maxwell body characterized by the equation (Maxwell 1868)

$$\frac{d\sigma}{dt} = E^* \frac{d\epsilon}{dt} - \frac{\sigma}{\tau}$$

where σ = stress, ϵ = strain, E^* = appropriate elastic constant and $\tau = \eta/E^*$ = relaxation time of the body.

The characteristics of a Maxwell body can be conveniently represented by a mechanical model consisting of a viscous element, possessing Newtonian properties, in series with an elastic element obeying Hooke's law. When applied to the case of shearing stresses, the viscous element represents the viscosity of the glass, whereas the elastic element equals its shear elasticity. Dynamic measurements when interpreted in terms of Maxwell elasticity and viscosity give the apparent values of these quantities at the particular frequency of the experiment. Only if the substance behaves exactly like the simple Maxwell body will the elasticity and viscosity be independent of frequency.

The following expressions were used to calculate the Maxwell shear elasticity μ and viscosity η_m from the measured quantities. At low frequencies when the stress, strain and phase angle between them were measured

$$\mu = \frac{2}{\pi} \frac{l}{a^4} \frac{S}{\cos\delta} \frac{1}{\theta}, \quad \eta_m = \frac{\mu \cot\delta}{2\pi\nu}$$

where S = stiffness of the torsion wire in dyn cm/radian, δ = phase angle between the stress and strain, θ = ratio of the strain to the stress, a(cm) = radius of the glass rod, l(cm) = length of the glass rod, ν(c/s) = frequency. When the frequency and logarithmic decrement Δ of a mechanical resonant system performing free oscillations were determined,

$$\mu = 8\pi\nu^2 Il/a^4, \quad \eta_m = \mu/2\Delta\nu.$$

In the case where an elastic element was added

$$\eta_m = \frac{4\pi I \nu_2 \nu_1 l}{(\Delta_2\nu_1 - \Delta_1\nu_2)a^4}$$

where ν_1, ν_2 are frequencies with the glass rod absent and present respectively, Δ_1, Δ_2 are logarithmic decrements with the glass rod absent and present, and I = moment of inertia of the disc.

At high audio frequencies when no added inertia was involved

$$\mu = \rho(\lambda\nu)^2, \quad \eta_m = \mu/2\Delta\nu$$

where ρ = density in g cm^{-3} and $\lambda = 4l$.

Dr. R. S. Marvin (private communication) has pointed out that results of experiments involving exponentially decaying oscillations do not correspond to measurements at a single frequency, since the damped oscillations represent a narrow spectrum of frequencies. The band-width $\delta\nu$ of the spectrum at half its peak amplitude is given by $\delta\nu/\nu = \Delta/\pi$ provided $\Delta \ll 1$.

A single measurement thus gives the average effect over a small range of frequencies, and provided Δ is less than unity, the error introduced is small, since the quantities measured are not highly dependent upon frequency and the Fourier spectrum of a damped exponential is symmetrical about its central frequency.

In the temperature ranges involved, the damping of the torsional oscillations was sufficiently large for the damping of the surrounding air to be neglected. The measured logarithmic decrement due to air damping was found to be negligibly small compared with that measured with the glasses.

§ 4. Discussion of Results

Preliminary measurements showed that the organic glasses employed behave perfectly linearly with respect to stress at the order of strains used. This has been repeatedly verified by the exponential decay curves obtained. In figure 2 the logarithm of the Maxwell viscosity is plotted against the logarithm of the frequency for hydroxy-pentamethylflavan at several different temperatures. The 'static' viscosities, measured by Crawford (1953), are shown on the left-hand side and the dynamic measurements have been extrapolated back to link up with these. Figure 3 shows a similar set of curves for glycerol sextol phthalate, with the static values for this particular sample also plotted. The elasticity results are plotted in figure 4 for hydroxy-pentamethylflavan; those for glycerol sextol phthalate are of a similar nature.

Referring to figures 2 and 3, since the dynamic viscosity and elasticity of hydroxy-pentamethylflavan are independent of frequency over a very wide range it follows that this material behaves as a simple Maxwell body over this range. As the frequency is increased the viscosity falls, showing that the material no longer behaves in a simple manner. As the temperature is increased the frequency range over which the viscosity is constant increases.

Figure 3 shows that the sample of glycerol sextol phthalate also behaves as a Maxwell body at low frequencies, but in this case the range is very much narrower than with hydroxy-pentamethylflavan. An alternative method of showing the departure in the behaviour of the glasses from that of a simple Maxwell body at high frequencies is to plot the logarithm of the logarithmic decrement against the logarithm of the frequency, as in figure 5. The variations of the logarithmic decrement with frequency of a typical Maxwell body and a Voigt body (i.e. an elastic and viscous element in parallel) are given in the same figure. Curves are shown for hydroxy-pentamethylflavan at 16°c, glycerol sextol phthalate at 26°c and polymethyl methacrylate at 20°c, the latter being calculated from Lethersich's published results for the viscosity and shear elasticity (Lethersich 1950). The logarithmic decrement of polymethyl methacrylate varies little with frequency compared with the variations which occur with organic glasses. At low frequencies the curves for both glasses are parallel to that of the Maxwell body, but as the frequency increases the behaviour of the glasses becomes progressively more similar to that of polymethyl methacrylate. Thus the curve for glycerol sextol phthalate is parallel to that of the Maxwell body up to a frequency of about

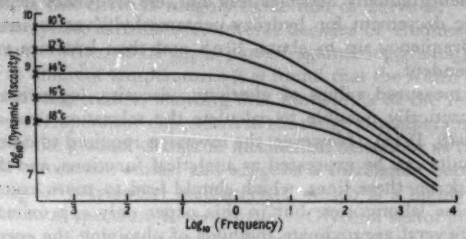


Figure 2. Dynamic viscosity of hydroxy-pentamethylflavan.

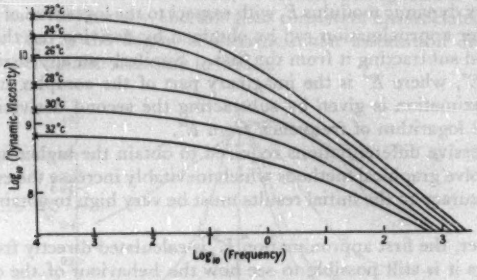


Figure 3. Dynamic viscosity of glycerol sextol phthalate.

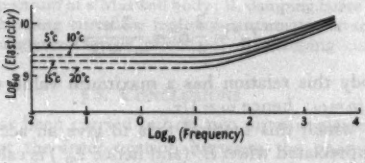


Figure 4. Dynamic elasticity of hydroxy-pentamethylflavan.

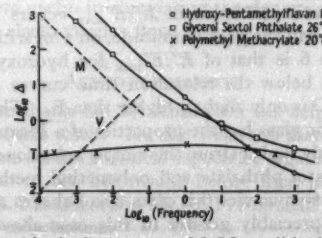


Figure 5. Logarithmic decrements. M, Maxwell body; V, Voigt body.

10^{-2} c/s and then gradually becomes less and less frequency dependent. Also, the logarithmic decrement for hydroxy-pentamethylflavan is inversely proportional to the frequency up to about 10 c/s and then becomes very much less frequency dependent.

From the measured values of elasticity, viscosity, logarithmic decrement, etc., it is theoretically possible to calculate the relaxation-time spectra (Kuhn 1947, Gross 1948, 1953). However, the inversion required to do this is tractable only if the results can be expressed as analytical functions, and this is generally difficult. Work on these lines, which should lead to more exact spectra, is in progress at these laboratories, but in this paper only approximate spectra are considered. Several approximate methods of obtaining the spectra have been proposed (Alfrey 1948, Andrews 1952, Schwarzl and Staverman 1952). It has been shown by these authors that the first approximation to the relaxation distribution function is proportional to the differential coefficient of the real part E' of the complex dynamic modulus E, with respect to the logarithm of the frequency. A higher order approximation can be obtained by deriving the third differential coefficient and subtracting it from the first. Similarly an approximate spectrum is given by E'', where E'' is the imaginary part of the complex modulus, and a second approximation is given by subtracting the second derivative of E'' with respect to the logarithm of frequency from E''.

The successive differentiations required to obtain the higher approximations generally involve graphical methods which inevitably increase the errors involved. Hence the accuracy of the initial results must be very high to obtain a satisfactory spectrum.

If, however, the first approximation E'' is calculated directly from the experimental results it is still possible to see how the behaviour of the organic glasses compares with that of other materials. In terms of the logarithmic decrement and the Maxwell elastic modulus, the value of the first approximation to the relaxation spectrum (L_0'' in Schwarzl's notation) is given by

$$L_0'' = \frac{2}{\pi} E'' = \frac{2\Delta\mu}{\pi^2 + \Delta^2}.$$

For a Maxwell body this relation has a maximum value when $\Delta = \pi$, or, since $\Delta = \mu/2\eta\nu$, when $\mu/\eta = \omega$; hence $\omega = 1/\tau$.

The extent by which this method fails to give an accurate relaxation time spectrum can be appreciated when E'' (and hence L_0'') is calculated for a Maxwell body. In figure 6 the true Maxwell relaxation spectrum is line A, and curve B is the plot of E''/E''_{max} against $\log_{10}(\omega/\omega_0)$, ω_0 being the value of ω when E''/E''_{max} is a maximum. The maximum of curve B coincides with A and hence for a Maxwell body the maximum value of E''/E''_{max} occurs at the relaxation time although the width, at half the peak amplitude, on a logarithmic scale is about 1·2.

Curve C of figure 6 is that of E''/E''_{max} for hydroxy-pentamethylflavan at 16°c. At frequencies below the relaxation time curves B and C coincide and at higher frequencies C is only slightly higher than B. Thus below the predominant relaxation time the glass has the properties of a Maxwell body and at higher frequencies it differs only slightly from this simple behaviour. The corresponding curves for glycerol sextol phthalate and polymethyl methacrylate are D and E, respectively. At low frequencies this glass also behaves as a Maxwell body, but the discrepancy is appreciably greater in this case above the relaxation time.

In fact the width of this curve is roughly twice that of curve B. Curve E, for polymethyl methacrylate, extends over several decades of frequency and has a very flat maximum. When E''/E''_{max} is plotted against $\log_{10}(\omega/\omega_0)$ for glycerol sextol phthalate at different temperatures it is found that the width of the curves changes little with temperature, showing that the widths of the relaxation-time spectra are not very temperature dependent in the range of temperatures involved.

Figure 6 shows the gradual change in mechanical behaviour as the possibility of intermolecular association varies, and suggests a correlation between the two properties. Thus for polymethyl methacrylate, which is a highly cross-linked polymer, the relaxation spectrum is very broad. Glycerol sextol phthalate has a relaxation spectrum which is considerably different from that of a Maxwell body, but not nearly as broad as that of polymethyl methacrylate. The chemical nature of this glass, as has already been stated, implies that there is likely to be considerable association between the molecules. Further, the spectrum of hydroxy-pentamethylflavan is very narrow, and this glass consists of unassociated molecules. Thus it appears that as the degree of intermolecular association decreases, the relaxation-time spectra also decrease in width.

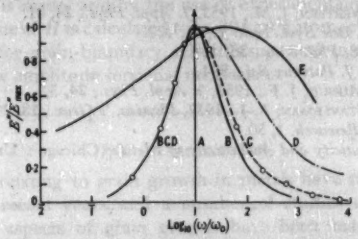


Figure 6. A, relaxation spectrum of a Maxwell body; B, damping curve of a Maxwell body; C, experimental damping curve for hydroxy-pentamethylflavan; D, experimental damping curve for glycerol sextol phthalate; E, damping curve for polymethyl methacrylate.

§ 5. Conclusions

The work which has been carried out on two organic glasses has shown how these materials suggest the inter-relation between the chemical and physical properties of amorphous substances. When other glasses have been studied it should be possible to establish the correlation in much greater detail. Also the extension of the measurements to higher frequencies and to different types of deformation should prove useful in elucidating the departure of the behaviour of the glasses from simple elasto-viscous flow.

Acknowledgments

The author wishes to thank Dr. H. Kolsky for his help and encouragement throughout the course of this work. He also wishes to thank Dr. D. J. C. Wood for his helpful discussions on the chemical aspects of organic glasses, Miss S. M. Crawford for permission to use her results before final publication, and Miss P. G. Bailey, Mr. A. F. L. Lassauzet and Mr. D. F. Marshall who carried out many of the experiments and calculations.

References

Alfrey, T., 1948, *Mechanical Behaviour of High Polymers* (New York : Interscience).
Andrews, R. D., 1952, *Industr. Engng Chem.* (*Industr.*), **44**, 707.
Baker, W., Curtis, R. F., and McOmie, J. F. W., 1952, *J. Chem. Soc.*, 1774.
Benbow, J. J., 1953, *J. Sci. Instrum.*, **30**, 412.
Crawford, S. M., 1953, *Proc. Phys. Soc.* B, **66**, 953.
Davies, R. M., and James, E. G., 1934, *Phil. Mag.*, **18**, 1023.
Ferry, J. D., 1953, *J. Colloid Sci.*, **8**, 224.
Fitzgerald, J. V., 1951, *J. Amer. Ceram. Soc.*, **34**, 314, 339, 388, 390.
Fox, F. E., and Litovitz, T. A., 1951, *Proceedings of the International Colloquium on Ultrasonic Vibrations, Brussels*, p. 38.
Gross, B., 1948, *J. Appl. Phys.*, **19**, 257; 1953, *Mathematical Structure of the Theory of Visco-elasticity* (Faris : Hermann), Ch. V.
Hillier, K. W., 1949, *Proc. Phys. Soc.* B, **62**, 701.
Jones, G. O., 1948–49, *Rep. Progr. Phys.*, **12**, 133 (London : Physical Society).
Kamel, R., 1949, *Phys. Rev.*, **75**, 1606.
Kuhn, W., 1947, *Helv. Chim. Acta*, **30**, 307, 464.
Lethersich, W., 1950, *Brit. J. Appl. Phys.*, **1**, 294.
Markovitz, H., 1952, *J. Appl. Phys.*, **23**, 1070.
Markovitz, H., Yavorsky, P. M., Harper, R. C., Zapas, L. J., and DeWitt, T. W., 1952, *Rev. Sci. Instrum.*, **23**, 430.
Marx, J. W., and Sivertsen, J. M., 1953, *J. Appl. Phys.*, **24**, 81.
Mason, W. P., 1951, *Bell Syst. Tech. J.*, **31**, 122.
Maxwell, J. C., 1868, *Phil. Mag.*, **35**, 129.
Nolle, A. W., 1950, *J. Polymer Sci.*, **5**, 1,
Nolle, A. W., and Mifsud, J. F., 1953, *J. Appl. Phys.*, **24**, 5.
Schwarzl, F., and Staverman, A. J., 1952, *Physica, 's Grav.*, **18**, 791.
Weyl, W. A., 1948, *Research*, **1**, 50.
Zener, C., 1948, *Elasticity and Anelasticity of Metals* (Chicago : University Press).

On the Mechanism of Grain Growth in Metals, with Special Reference to Steel

By D. G. COLE, P. FELTHAM and E. GILLAM
Physics Laboratories,
The British Iron and Steel Research Association, London

Communicated by N. F. Mott; MS. received 26th October 1953

Abstract. The rate of grain growth in the initial stages of isothermal austenitizing has been studied by a thermal etching method in plain, eutectoid, carbon steel in the range 840–970°C. An equation relating the mean initial and instantaneous grain diameters, D_0 and D, to the austenitizing time t, of the form $D^2-D_0^2=Kt\exp(-H/\mathbf{k}T)$ is developed on the assumption that the growth rate is controlled by interfacial tension at grain boundaries. An estimate of the constant K is made, and by the use of experimentally obtained values of D, D_0 and t the value of H is calculated to be 30 kcal/g atom, close to the activation energy estimated for grain-boundary self-diffusion and to that determined for the damping of low amplitude torsional oscillations.

§ 1. Introduction

Problems relating to grain growth in metals have received considerable attention in recent years, and a number of studies of the practical and fundamental aspects of grain growth have been made. Beck, Kremer, Demer and Holzworth (1947) studied grain-growth in aluminium and in aluminium–magnesium alloys, Burke (1948) and Beck, Towers and Manley (1948) in brass, Cleaves and Hiegel (1942) in gamma-iron, and Shushpanov (1937), Day and Austin (1940), Whiteley (1943), Gudtsov and co-workers (1950), Miller (1951) and Feltham (1953) in austenite.

A number of fundamental papers on grains and grain growth have recently been published by C. Smith (1952, 1953), and the subject has been reviewed by Burke and Turnbull (1952). Burke (1948) in discussing the experimental evidence relating to grain growth concluded that this supported the belief that the interfacial tension at grain boundaries determined the net rate of transfer of atoms from one grain to another. His own results on the grain growth in high-purity brass were reconcilable with the assumption that, with certain qualifications relating to inclusions, the rate of boundary migration at a given temperature was inversely proportional to the radius of curvature of the boundaries, and that this radius was generally of the same order of magnitude as the mean grain diameter of the sample. Hence from the relation $dD/dt=A/D$, Burke obtained the relation $D^2-D_0^2=Kt\exp(-H/\mathbf{k}T)$ where K is a constant. He did not investigate the nature of the constant and of the activation energy, but as the growth mechanism must be one involving the transfer of atoms at grain boundaries from one grain to another, the activation energy would be expected to be approximately equal to that of grain boundary self-diffusion

The principal aims of the present work were to obtain data on the grain growth of steel which would lend themselves to a study of the mechanism of the process and, more particularly, to examine the validity of Burke's simple relation during the rapid, initial, growth occurring during austenitizing, to attempt an evaluation of the constant K experimentally, and to compare this value with one roughly estimated on the basis of the assumed mechanism of growth.

§2. Experimental

In each experiment a strip of steel of composition 0·79% C, 0·15% Si, 0·54% Mn, 0·04% S, 0·04% P, 0·03% Cr, 0·04% Ni and 0·03% Cu, 3 cm long, 0·5 cm wide and 0·02 cm thick, was mechanically polished and subsequently heated in a continuously evacuated vacuum chamber at a pressure of about 0·2 microns Hg of air. The temperature was measured by means of a chromel–alumel thermocouple which was welded to the specimen.

The strips were annealed for a few minutes at 400°C, and then raised to the austenitizing temperature (840°, 900°, 950° or 970°C respectively) at which they were allowed to undergo grain growth for periods ranging from 1·5 to 30 minutes. After each heat treatment of this type the temperature was dropped

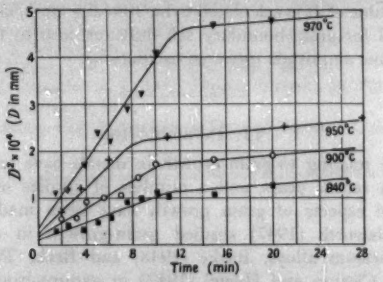


Figure 1. Grain growth in plain 0·79% C steel.

rapidly to 580°C by decreasing the heating current flowing through the specimen. At that temperature the austenite transformed to pearlite within about one minute. The current was then switched off so that the specimen cooled to room temperature within about a further minute. Thermal etching took place at the austenitizing temperature, and the austenite grain boundaries were consequently clearly visible on the surface of the cold specimens when examined by means of a microscope. The mean grain diameter of the austenite grains was measured by taking the square root of the average grain area over about 1000 surface grains, and also by the line-intercept method in some cases, though this was eventually discontinued. (Smith and Guttman (1953) have recently discussed the second method and its relation to others.) The curves relating D^2, the square of the instantaneous value of the mean grain diameter, and the time of isothermal grain-growth at four austenitizing temperatures are shown in figure 1. It can be seen that the initial portions can quite well be represented by straight lines. However, at a certain grain size, which depends upon the temperature at which isothermal growth has taken place (Feltham 1953) the mean growth rate falls to a very small value which does not appear to be significantly temperature dependent within the range of temperatures used. The probable origin of this

sharp decline in the growth rate, which is not observed in the case of the growth of foam cells (Smith 1952) will be considered below, in connection with the discussion of the mechanism of growth in metal polycrystals.

§ 3. Theoretical

Consider a diameter D_1 in a given grain, growing by the addition of atoms to both ends. From the point of view of the accuracy aimed at in the present treatment it is sufficient to regard the atoms of the lattice as Einstein oscillators capable of vibrating in any plane about their respective mean positions with a single frequency ν, which we take to be approximately equal to $\boldsymbol{kT/h}$ (Glasstone, Laidler and Eyring 1941, Nowick and Machlin 1947). In order that an atom shall be able to jump from the adjacent grain to the end of the diameter D_1 the amplitude vector of the oscillation must lie within a cone which, in a close-packed cubic lattice, would contain only about four of its twelve nearest neighbours, so that the frequency of oscillations capable of leading to atom transfers across the grain boundary is ν/m, where m is about 3. The diameter D_1 will therefore grow at the rate

$$\frac{dD_1}{dt} = 2\frac{\nu}{m}ap = \frac{2}{3}\frac{\boldsymbol{kT}}{\boldsymbol{h}}ap \qquad \ldots\ldots(1)$$

where a is the interatomic spacing and p the probability of the net transfer of an atom to one end of the diameter in the time m/ν.

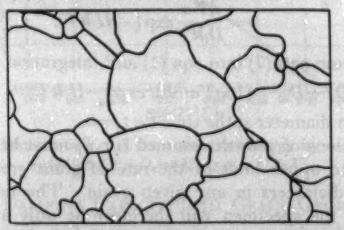


Figure 2. Austenite grains towards the end of the period of rapid grain growth at 950 °c (×390). (Transcribed from micrograph.)

The assumption is made that the curvature of the boundary surfaces at the ends of the diameter is negative, so that the grain is doubly concave, and (Burke 1948) that both principal radii of curvature are equal to D_1. It can be seen from figure 2, showing grains towards the end of the period of rapid growth, that the last assumption is plausible. However, in very pure metals and in soap films a structure of almost equi-axed polyhedra tends to be approached very quickly, and the radii of curvature will, in general, be considerably larger than the grain diameters. In the ideal case which, contrary to the belief sometimes expressed (viz. Harker and Parker 1945), is unattainable in practice, except perhaps in bi-crystals, the boundaries are ultimately flat and the ratio of the radius of curvature to that of the grain diameter infinite. The assumption that the ratio is equal to unity is nevertheless likely to be approximately true only up to a certain period of grain growth. In the present case this period extends over about eleven minutes (figure 1), to which the validity of the following theoretical

considerations is restricted. Subsequent growth takes place at a comparatively slow rate which does not appear to be sensibly temperature dependent.

Now if an element of this concave grain surface of area dA advances a distance dD_i normal to itself, it changes in area by $(2/D_i)dAdD_i$, while it sweeps out a volume $dAdD_i$. Hence the mean change in surface energy per atom of volume V, involved in the advance, is equal to

$$W = 2V\sigma/D_i \qquad \ldots\ldots(2)$$

where σ is the specific surface energy.

Thus the activation energy for the jump of a boundary atom to the diameter D_i is

$$Q_1 = H - W, \qquad \ldots\ldots(3)$$

and for a jump away from the diameter

$$Q_2 = H + W, \qquad \ldots\ldots(4)$$

where H is the activation energy (enthalpy) of boundary self-diffusion.

Writing

$$p = \exp\{-(H-W)/\boldsymbol{k}T\} - \exp\{-(H+W)/\boldsymbol{k}T\} \qquad \ldots\ldots(5)$$

or, since $W \ll \boldsymbol{k}T$,

$$p = \frac{2W}{\boldsymbol{k}T}\exp\{-H/\boldsymbol{k}T\}, \qquad \ldots\ldots(6)$$

whence, substituting for W from eqn (2),

$$p = \frac{4V\sigma}{D_i\boldsymbol{k}T}\exp(-H/\boldsymbol{k}T). \qquad \ldots\ldots(7)$$

Substitution for p from eqn (7) into eqn (1) and integration yields

$$D_i^2 - D_{i0}^2 = (16aV\sigma/3\boldsymbol{h})t\exp(-H/\boldsymbol{k}T), \qquad \ldots\ldots(8)$$

where D_{i0} is the grain diameter at the time $t = t_0 = 0$.

The ideal conditions of growth assumed for D_i must be regarded as leading to approximately the upper limit of the rate of grain growth, which will not be attained for all diameters in any given grain. The rate of growth of the mean grain area of the specimen will therefore be only a fraction of dD_i^2/dt. Statistical data from which an estimate of this fraction could be made are not available, and it will be assumed that 0·1 represents a reasonable guess, the fraction being regarded as independent of the grain size and the temperature.

If D is the mean grain diameter of the sample it then follows that

$$D^2 - D_0^2 = (16aV\sigma/30\boldsymbol{h})t\exp(-H/\boldsymbol{k}T). \qquad \ldots\ldots(9)$$

On substituting into the bracketed term in eqn (9) the values for iron: $\sigma = 2\times10^3$ dyn cm^{-1}, $V = 1{\cdot}2\times10^{-23}$ cm^3, $a = 3\times10^{-8}$ cm, and $\boldsymbol{h} = 6{\cdot}5\times10^{-27}$ ergsec, one obtains

$$D^2 - D_0^2 = Kt\exp(-H/\boldsymbol{k}T) \qquad \ldots\ldots(10)$$

where

$$K = 6\times10^{-2}\ \text{cm}^2\,\text{sec}^{-1}, \qquad \ldots\ldots(11)$$

a constant having the dimensions of a coefficient of diffusion.

No experimentally determined value of H is available, but on the assumption that it is equal, or close, to the activation energy of grain-boundary self-diffusion, an estimate of its magnitude can be made. The basis of this estimate lies in the fact that grain boundary self-diffusion should have an activation energy larger than that for surface self-diffusion but smaller than that for volume (single crystal)

self-diffusion, and that the apparent (polycrystal) activation energy for self-diffusion and that for volume self-diffusion should be approximately equal at high temperatures (Le Claire 1953). Data on the ratio of the activation energy for boundary diffusion to that for the volume or apparent self-diffusion given by Smithells (1949) and Le Claire (1953) yield the following values: Ag into Ag 0·5 to 0·8, W into W 0·74 ; Th into W 0·78; Ag into Cu 0·63. The value for the diffusion of zinc into copper (0·72) is that quoted by Smoluchowski (1952) in his paper " The Theory of Grain Boundary Diffusion". Taking 0·70 as the mean, and 42 kcal/g atom as the activation energy for volume self-diffusion of the steel (Gruzin *et al.* 1951), the estimated value of the activation energy H is 29·5 kcal/g atom, in good agreement with the value of 30 kcal/g atom obtained from the slope of the straight line relating $\log[(D^2-D_0^2)/t]$ to T^{-1}, based on eqn (1) (figure 3). The experimental value of K obtained by using $H=30$ kcal/g atom and smoothed values of T^{-1} and $\log[(D^2-D_0^2)/t]$, taken from figure 3, is found to be 5×10^{-2} cm^2 sec^{-1}, also in very good agreement with the theoretical value (eqn (11)).

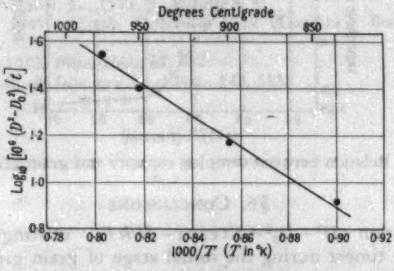


Figure 3. Determination of the activation energy by means of the equation $D^2-D_0^2=Kt\exp(-H/kT)$.

§4. Damping Capacity and Grain Growth

Cole (1954) has reported some experiments on the damping capacity of austenite in which, at a given temperature, he finds a variation of the logarithmic decrement with time, which is very similar to the variation of the grain size with time. Figure 4 shows how by an adjustment of ordinates both the logarithmic decrement, taken from values given by Cole, and the reciprocal of the grain diameter can be made to fall on to the same curve when plotted against $(t+t')^{-1/3}$. The constant t' is incorporated into the bracket to allow for the fact that D^2 is not zero at the time $t=0$. By extrapolating the grain growth curve t' is found to be approximately equal to two minutes. Using a method due to Kê (1947), Cole finds a value of about 35 kcal/g atom for the activation energy relating to the grain-boundary processes responsible for the damping, which is of the same order as the value derived here for the activation energy for boundary self-diffusion.

§5. Discussion

In view of the simplifying assumptions made in the present theoretical treatment of the initial grain growth, the good agreement between the theoretical and experimental results is certainly to some extent fortuitous, and further

work, particularly with metals of high purity, appears to be desirable. In aluminium of high purity the relation between grain size and time during the stage of initial growth was found to obey a relation of the type (Beck, Holzworth and Sperry 1948)

$$D^n - D_0^n \propto t \qquad \ldots\ldots(12)$$

with $n = 2{\cdot}3$ at 650°C, i.e. close to the value of 2 used throughout the present work. However, higher values were recorded at lower temperatures. Also, little is known about the nature of the sometimes pronounced effects of dissolved and dispersed impurities on the activation energies and coefficients of diffusion (Gruzin *et al.* 1951, Hoffman and Turnbull 1952), effects which undoubtedly will have to be taken into consideration in future work on grain growth.

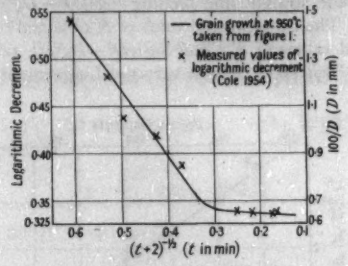


Figure 4. Relation between damping capacity and grain growth at 950°C.

§6. Conclusions

1. The equation $D^2 - D_0^2 = Kt\exp(-H/kT)$, relating the mean grain diameter and the time t during the initial stage of grain growth, expresses the latter well in a plain, eutectoid, steel in the range 840–970°C.

2. Close agreement is found between the experimentally determined value of K, and that determined theoretically on the basis of the assumption that the net rate of transfer of atoms during grain growth from one grain to another is determined by interfacial tensions.

3. The experimentally determined value of H (30 kcal/g atom) lies close to the value of the activation energy for grain boundary self-diffusion (29·5 kcal/g atom) estimated from the activation energy for volume self-diffusion given in the literature.

4. Grain boundary processes responsible for the damping of torsional oscillations at high temperatures appear to be related to those occurring in grain growth, the activation energies being approximately 35 and 30 kcal/g atom respectively.

References

Beck, P. A., Holzworth, M. L., and Sperry, P. R., 1948, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **175**, 372.

Beck, P. A., Kremer, J. C., Demer, L. V., and Holzworth, M. L., 1947, *Metals Tech.*, Sept. issue.

Beck, P. A., Towers, J., and Manley, W. D., 1948, *Metals Tech.*, Feb. issue.

Burke, J. E., 1948, *Metals Tech.*, Oct. issue.

Burke, J. E., and Turnbull, D., 1952, *Progress in Metal Physics* (London: Pergamon Press).

CLEAVES, H. E., and HIEGEL, J. M., 1942, *J. Res. Nat. Bur. Stand.*, **28**, 643.
COLE, D. G., 1954, *J. Iron and Steel Inst.*, in the press.
DAY, M. J., and AUSTIN, J. B., 1940, *Trans. Amer. Soc. Metals*, **28**, 643.
FELTHAM, P., 1953, *Proc. Phys. Soc.* B, **66**, 865.
GLASSTONE, S., LAIDLER, K. J., and EYRING, H., 1941, *The Theory of Rate Processes* (New York: McGraw-Hill).
GRUZIN, P. L., KORNEV, Y. V., and KURDYUMOV, G. V., 1951, *C. R. Acad. Sci., U.R.S.S.*, **80**, 49.
GUDTSOV, N. T., LOZINSKII, M. G., ZUDIN, F. I., BOGDANOV, N. A., and MATVEEVA, M. P., 1950, *Bull. Acad. Sci., U.R.S.S.*, (O.T.N.), **14**, 108.
HARKER, D., and PARKER, E., 1945, *Trans. Amer. Soc. Metals*, **34**, 156.
HOFFMAN, R. E., TURNBULL, D., 1952, *J. Appl. Phys.*, **23**, 1409.
KÊ, T. S., 1947, *Phys. Rev.*, **72**, 41.
LE CLAIRE, A. D., 1953, *Progress in Metal Physics*, Vol. 4 (London: Pergamon Press).
MILLER, O. O., 1951, *Trans. Amer. Soc. Metals*, **43**, 260.
NOWICK, A. S., and MACHLIN, E. S., 1947, *J. Appl. Phys.*, **18**, 79.
SHUSHPANOV, L. I., 1937, *Metallurg*, **12**, 31.
SMITH, C., 1952, *Metal Interface* (Cleveland: American Society of Metals), p. 65; 1953 *Trans. Amer. Soc. Metals*, **45**, 533.
SMITH, C., and GUTTMANN, L., 1953, *J. Metals*, **5**, 81.
SMITHELLS, C. J., 1949, *Metals Reference Book* (London: Butterworths Scientific Publications).
SMOLUCHOWSKI, R., 1952, *Phys. Rev.*, **87**, 482.
WHITELEY, J. H., 1943, *J. Iron and Steel Inst.*, **148**, 513.

The Determination of the Refractive Indices of Inhomogeneous Solids by Interference Microscopy *

By R. C. FAUST

The British Rayon Research Association, Urmston, near Manchester

MS. received 9th September 1953

Abstract. The refractive indices of an optically inhomogeneous specimen will generally vary in directions both parallel to and perpendicular to the microscope axis. It is, however, possible to define a mean index μ_s such that ($\mu_s \times$ specimen thickness) represents the optical thickness in a direction parallel to this axis. Interference microscopy reveals those parts of the specimen for which the mean index μ_s is equal to the refractive index of the immersion liquid. Thus, by gradually changing the index of the liquid relative to that of the specimen, one can determine the manner in which μ_s varies from region to region of the specimen. These changes in relative indices can be effected by altering either the wavelength of the incident light or the temperature of the system. In the wavelength variation method, which is applicable to specimens of thickness less than about 0·1 mm, the incident beam is of white light, and the fringe pattern is observed with the aid of a spectrograph. In the temperature variation method, which is useful for thicker specimens, monochromatic light is employed, the fringes being observed by direct microscopic examination.

If the specimen is of convenient geometrical form it is possible to utilize these observed values of μ_s in order to calculate the refractive indices at any point in the interior of the specimen, but with irregularly shaped specimens one might have to resort to section cutting in order to achieve the same results. Application of these methods to anisotropic bodies enables conclusions to be drawn concerning the degree of molecular orientation and the closeness of molecular packing at different regions of the specimen.

§1. Introduction

One of the aims of the worker in the field of textile physics is to relate the structural arrangement of the molecules within a fibre to the mechanical properties of the fibre and, in the case of man-made fibres, to the mode of production. This problem is complicated by the fact that most fibres are not homogeneous structures. Of the natural textile fibres reference may be made to cotton, a plant cell consisting of a lumen surrounded by a secondary and a primary wall, and to wool, an animal protein fibre with an outer scaly structure. Such inhomogeneity of structure is also manifest in many regenerated cellulose (rayon) fibres, the molecules in the outer regions generally being more highly oriented than those in the underlying core. Similar structural variations might also occur in such synthetic melt-spun fibres as nylon and terylene.

The elucidation of these structures would be greatly facilitated if the birefringence were known for each region of the specimen, for such information

* This work forms part of the fundamental research programme undertaken by the British Rayon Research Association.

would provide a measure of the degree of scatter of the molecules about the preferred direction of orientation. However, even more information can be gained if, instead of the birefringence, the absolute refractive indices of the specimen are measured. By applying the empirical relationship of Gladstone and Dale to anisotropic bodies, Hermans (1949) has shown that the density d of a given substance is related to its optical properties by the expression $(\mu_{iso}-1)/d=$constant, where μ_{iso} is the arithmetic mean of the three principal refractive indices. (The quantity μ_{iso} can be termed the effective isotropic refractive index, since it is the index the specimen would have if the orientation were destroyed without altering the density.) Thus, if the principal refractive indices are determined for all parts of the specimen, one can follow changes in both the birefringence (molecular orientation) and the effective isotropic index (density or closeness of packing of the molecules).

Since the conventional methods of immersion refractometry are unsuited to the examination of optically inhomogeneous specimens (Faust 1952), it has been found necessary to employ interference microscopy for observing spatial changes in the refractive indices of such specimens. As with all immersion methods, the index of the liquid must be changed relative to that of the solid, and this has been achieved by altering either the wavelength of the incident light or the temperature of the system. Although these methods are illustrated by their application to textile fibres, they can be used in a wide variety of studies. In biological research, for instance, it would be possible to investigate not only cylindrical cells such as nerve fibres but also cells of any other shape. The study of living cells might, however, present some difficulty because aqueous protein solutions must be used as immersion media (Barer, Ross and Tkaczyk 1953). The properties of the immersion medium and the living cell might then be so similar that a change of wavelength or temperature would bring about only a very slight alteration in the relative indices. This difficulty could perhaps be overcome by gradually altering the concentration of the protein solution.

§ 2. Wavelength Variation Method

This method has already been described in detail (Faust 1952). The solid specimen, immersed in a liquid of approximately the same refractive index, is placed between two parallel optical flats, the inner surfaces of which have been coated with highly reflecting, partially transmitting silver films. When this interferometer is illuminated at normal incidence by a parallel beam of white light (figure 1 (*a*)), it acts as a wavelength filter, passing certain wavelengths and suppressing others. If an image of the liquid-immersed specimen is projected on to the slit of a spectrograph, the transmitted wavelengths can be observed as bright fringes. The wavelength λ_n of the nth order fringe satisfies the relationship

$$(n-\Omega_n)\lambda_n=2(\mu_L)_n t_g+2(\mu_S-\mu_L)_n t \qquad \ldots\ldots(1)$$

where t_g is the constant interferometer gap thickness, t the specimen thickness, and Ω_n the total phase change, expressed as a fraction of the wavelength, that occurs upon reflection at the two silver films. The refractive indices $(\mu_L)_n$ and $(\mu_S)_n$ are those of the liquid and the solid at the wavelength λ_n (in general the refractive index of the solid will not be constant along a line parallel to the microscopic axis, and $(\mu_S)_n$ will therefore represent a mean index so defined that the optical thickness of the solid is given by $(\mu_S)_n t$).

In those regions of the interferometer containing liquid only ($t=0$), the wavelength λ_n of a given fringe is constant. Hence, in the absence of astigmatic curvature in the spectrograph, each fringe will be a straight line at right angles to the wavelength axis (figure 1 (*b*)). Upon crossing into a region containing both liquid and solid, each fringe will in general undergo a wavelength displacement. This displacement will, however, vanish whenever the index of the liquid equals that of the solid. It is thus seen from figure 1 (*b*) that $(\mu_S)_n=(\mu_L)_n$ at the points P_n of the object, whilst $(\mu_S)_{n+1}=(\mu_L)_{n+1}$ at the points P_{n+1}, and so on. The indices $(\mu_L)_n$, $(\mu_L)_{n+1}\ldots$. corresponding to the wavelengths λ_n, $\lambda_{n+1}\ldots$. can be readily obtained by using an Abbe or Pulfrich refractometer to determine the dispersion curve of the liquid. If the experiment is now repeated using a liquid of slightly different refractive index, the indices at the points P_n, $P_{n+1}\ldots$. of the solid can be determined for the new wavelengths λ^*_n, $\lambda^*_{n+1}\ldots$. In this way the dispersion curve of the solid can be constructed

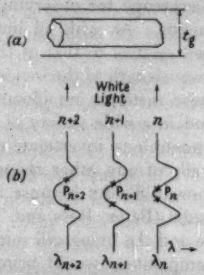


Figure 1. Wavelength variation method: (*a*) parallel plate interferometer; (*b*) fringes formed by projecting an image of a fibre on to a spectrograph slit, the fibre axis and slit being at right angles. Because of the irregular shape of the fibre, the fringes will generally be asymmetrical about the fibre axis.

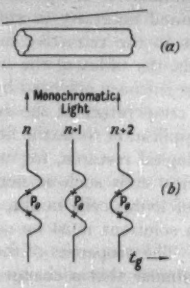


Figure 2. Temperature variation method: (*a*) wedge interferometer; (*b*) fringes formed by a fibre with its axis along the line of greatest slope of the wedge.

for each of the points P_n, $P_{n+1}\ldots$., and from these can be obtained a plot of refractive index across the solid for any desired wavelength λ. This procedure in no way involves a knowledge of the specimen thickness.

A simplification arises when, as is often true, the variation of index across the specimen is small. It is then reasonable to assume that the dispersion will be the same for all parts of the solid, and this can be measured by the relatively rapid central illumination (Becke line) method. Hence the interferometric determination need be done with only one immersion liquid.

§3. Temperature Variation Method

3.1. *Principle*

In the wavelength variation method the wave-number separation between adjacent fringes is inversely proportional to the optical thickness of the interferometer gap, and with specimen thicknesses greater than about 0·1 mm. the fringes are so close as to appear as a continuous spectrum. With such

specimens it is therefore necessary to restrict the incident light to monochromatic radiation (mercury green line), thereby eliminating the possibility of using wavelength variations in order to alter the index of the liquid relative to that of the solid. This disadvantage is overcome by changing the temperature of the system, the temperature coefficient of refractive index for a liquid generally being about ten times as large as that for a solid.

An interferometer with a wedge-shaped gap is illuminated at normal incidence by a parallel beam of monochromatic light, the resultant fringe system being viewed through a microscope directly (figure 2 (*a*)). These are fringes of equal optical thickness, to which eqn (1) now applies in the modified form

$$(n-\Omega)\lambda = 2(\mu_L)_\theta t_g + 2(\mu_S - \mu_L)_\theta t. \qquad \ldots\ldots(2)$$

Here λ is the wavelength of the monochromatic light, Ω the phase term for this wavelength, n the fringe order, t_g the variable gap thickness, t the specimen thickness, and $(\mu_L)_\theta$ and $(\mu_S)_\theta$ the refractive indices for the liquid and the solid for the wavelength λ and the temperature θ. In those parts of the interferometer containing only liquid ($t=0$), the fringes are straight lines parallel to the wedge apex. Upon entering the specimen region the fringes will generally undergo a linear displacement dx_θ (figure 2 (*b*)). This will correspond to a change in the gap thickness t_g of $\frac{1}{2}\lambda(dx/\Delta x)_\theta/(\mu_L)_\theta$, where Δx_θ is the linear separation of adjacent fringes. Combined with eqn (2), this yields the relationship

$$\tfrac{1}{2}\lambda(dx/\Delta x)_\theta = (\mu_L - \mu_S)_\theta t, \qquad \ldots\ldots(3)$$

$(dx/\Delta x)_\theta$ being positive if the displacement is away from the wedge apex. The displacement vanishes only at those points, such as P_θ, for which $(\mu_L)_\theta = (\mu_S)_\theta$. If the temperature is now changed to θ_1, points P_{θ_1} can be similarly discovered for which $(\mu_L)_{\theta_1} = (\mu_S)_{\theta_1}$. Since $(\mu_S)_\theta = (\mu_L)_\theta$ for the points P_θ of the specimen, the index for the same points at the arbitrary temperature θ_0 is

$$(\mu_S)_{\theta_0} \text{ at } P_\theta = (\mu_L)_{\theta_0} - [(\mu_S)_\theta - (\mu_S)_{\theta_0}]. \qquad \ldots\ldots(4)$$

In the following section it is shown how this equation can be used to determine the refractive index at any point of the solid for the fixed wavelength λ and a fixed temperature θ_0. As with the wavelength variation method, it is unnecessary to know the specimen thickness.

3.2. *Determination of the Index at Temperature* θ_0

A standard object (e.g. a right cylinder) of known thickness and suitable refractive index μ_S' is placed between the interferometer plates alongside the specimen. If the temperature of the system is changed from θ_0 to θ the fringe displacement, as measured for the standard object, undergoes a change given by

$$(\mu_L - \mu_S')_\theta - (\mu_L - \mu_S')_{\theta_0} = \frac{\lambda}{2t'}\left[\left(\frac{dx}{\Delta x}\right)_\theta - \left(\frac{dx}{\Delta x}\right)_{\theta_0}\right]. \qquad \ldots\ldots(5)$$

Combined with eqn (4) this leads to

$$(\mu_S)_{\theta_0} \text{ at } P_\theta = (\mu_L)_{\theta_0} + \frac{\lambda}{2t'}\left[\left(\frac{dx}{\Delta x}\right)_\theta - \left(\frac{dx}{\Delta x}\right)_{\theta_0}\right] + [(\mu_S')_\theta - (\mu_S')_{\theta_0}] + [(\mu_S)_\theta - (\mu_S)_{\theta_0}]. \qquad \ldots\ldots(6)$$

If the standard object is composed of the same type of material as the specimen, the temperature coefficients $d\mu_S/d\theta$ and $d\mu_S'/d\theta$ can be equated with little error, and the last two terms of eqn (5) will then disappear. (A special example of

this circumstance arises when it is possible to measure the thickness of some part of the specimen; the standard object is then the specimen itself, the thickness t' now corresponding to the known specimen thickness.) To calculate $(\mu_S)_{\theta_0}$ it is now only necessary to determine $(\mu_L)_{\theta_0}$, $(dx/\Delta x)_{\theta_0}$ and $(dx/\Delta x)_{\theta}$. As θ_0 is the only temperature at which it is necessary to measure both the index of the liquid and the fringe displacement, it is convenient to take the room temperature as θ_0. The temperature of the system is then altered, and the displacement $(dx/\Delta x)_{\theta}$ is measured at the same time as the points P_θ are discovered.

If, on the other hand, $(d\mu_S/d\theta) \neq (d\mu_S'/d\theta)$, a correction must be added to the value of $(\mu_S)_{\theta_0}$ obtained in the manner described in the previous paragraph. As the index of the liquid changes far more rapidly than that of the solid, the approximation

$$\left(\frac{d\mu_L}{d\theta}\right)(\theta-\theta_0) \simeq \frac{\lambda}{2t'}\left[\left(\frac{dx}{\Delta x}\right)_{\theta} - \left(\frac{dx}{\Delta x}\right)_{\theta_0}\right] \qquad \ldots\ldots(7)$$

may be employed. Thus, by taking a rough value for $d\mu_L/d\theta$, the approximate temperature change $\theta-\theta_0$ can be calculated. In the textile fibres so far examined by the writer, the maximum variation in refractive index across the fibre did not exceed 0·002. As the majority of liquids within the index range 1·40 to 1·70 possess a temperature coefficient $d\mu_L/d\theta$ of about -5×10^{-4} per centigrade degree, an index variation of 0·002 is bridged by a temperature change $\theta-\theta_0$ of only 4° or 5°C. The last two terms of eqn (5) can be now combined into the single term $(d\mu_S'/d\theta - d\mu_S/d\theta)(\theta-\theta_0)$, which can be estimated provided that these two temperature coefficients are known. The temperature coefficient of a solid may be either positive or negative, but its magnitude is generally less than 5×10^{-5} per centigrade degree. Consequently these coefficients, which can be determined by the Becke line method (Frey-Wyssling and Wuhrmann 1939), need not be known accurately in order to calculate a correction term amounting to $\pm$ 0·0002 or even less.

An alternative procedure is to measure the temperature θ of the immersion liquid by inserting a thermocouple between the interferometer flats; an Abbe or Pulfrich refractometer can then be used to find the index $(\mu_L)_\theta$ of the liquid for the wavelength λ (5461 Å). In this way it is possible to obtain $(\mu_S)_{\theta_0}$ from eqn (4) by either neglecting or estimating the small term $[(\mu_S)_\theta - (\mu_S)_{\theta_0}]$.

§4. Interferometer Design

The interferometer shown in figure 3 is used for both the wavelength and the temperature variation methods. The lower optical flat F_1 is held in a located ring R_1. The upper flat F_2 is fitted in a ring R_2 supported by three pins P, the lower end of each pin resting on the inclined face of a wedge-shaped block W. By means of a fine screw S each block can be moved parallel to the line of greatest slope, thereby causing the corresponding pin to rise or fall in such a way as to provide a sensitive control of the position and tilt of F_2 with respect to F_1. The interferometer is effectively enclosed in an air-tight box, the flats F_1 and F_2 forming parts of the floor and roof respectively. A flange from the detachable roof fits into a mercury-filled channel C cut in the ring R_2, thus forming an air-tight seal which, being frictionless, does not exert a restraint upon the delicate adjustment of the flat F_2. Inside the box is a heating coil H for raising the temperature of the system. The temperature can be lowered by passing air,

cooled in a freezing trap, through the box by way of the two taps T, which also act as terminals for the heating coil. Furthermore, provision is made for the introduction of a thermocouple junction.

When the microscope objective has a focal length less than 16 mm, the working distance is too small to accommodate the flat F_2, which must be replaced by a thin glass coverslip mounted in a ring of similar design to R_2.

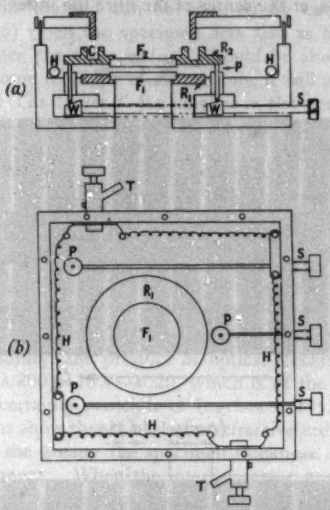


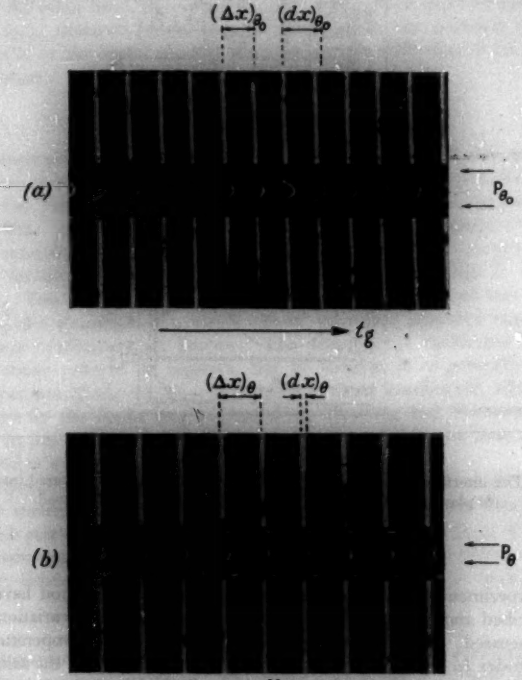
Figure 3. The interferometer : (*a*) elevation showing the lid being lowered into position; (*b*) plan; the lid, ring R_2 and the flat F_2 have been removed.

§ 5. Experimental Procedure

The experimental details of the wavelength variation method have already been described and only those pertaining to the temperature variation method will be discussed. At the wavelength 5461 Å and the room temperature θ_0 the refractive index $(\mu_L)_{\theta_0}$ of the immersion liquid should be nearly the same as that of the solid. Such a liquid is found from the Becke line procedure, the index $(\mu_L)_{\theta_0}$ being determined on an Abbe or Pulfrich refractometer. The solid and some of this liquid are placed between the silvered plates of an interferometer which, being illuminated at normal incidence by a parallel beam of mercury green light, is then adjusted to give satisfactory wedge fringes. If a standard object of known thickness is required, this too must be placed between the interferometer plates.

A model filament of viscose rayon has been used to illustrate this method; fortunately this fibre is of circular cross section (diameter 0·322 mm) and can therefore be used as the standard object. In figure 4 (*a*) is seen the fringe pattern observed at the room temperature θ_0. (The fibre acts as a uniaxial body with the

optic axis parallel to the fibre axis. Two plane-polarized fringe patterns are therefore produced, the vibration directions being parallel to and perpendicular to this axis. The fringes shown apply to the perpendicular vibrations.) The displacement $(dx/\Delta x)_{\theta_0}$ on the axis is 1·221; the sign is positive because the displacement is towards regions of greater gap thickness t_g, this being ascertained by opening the pinhole of the collimating system whereupon the fringes will broaden on that side corresponding to increasing gap thickness (Brossel 1946). Thus, at the centre of the fibre the index of the liquid is greater



× 30

Figure 4.

than that of the fibre, whilst at the edges the reverse is true (cf. eqn (3)). There is an intermediate position P_{θ_0} close to the fibre edge where the displacement vanishes, thereby indicating equality of the refractive indices of the liquid and the fibre, i.e. $(\mu_S)_{\theta_0} = (\mu_L)_{\theta_0}$ at P_{θ_0}. The temperature of the system is now altered; in this instance it is raised slightly so as to reduce the index of the liquid relative to that of the fibre. Figure 4 (*b*) shows the fringe pattern at the elevated temperature θ. (The different fringe spacings in the two photographs is due to a change in the wedge angle between the interferometer plates.) At the edge

of the fibre the fringes shoot towards the left by about an order and then move towards the right as the centre is approached, the position P_0 of zero displacement now being closer to the fibre axis than in figure 4 (*a*). The new axial displacement $(dx/\Delta x)_0$ is 0·138 which, when substituted in eqn (6), leads to the conclusion that at P_0 the index $(\mu_S)_{\theta_0} = (\mu_L)_{\theta_0} - 0{\cdot}0009_2$. By repeating this procedure and obtaining pairs of values for P_0 and $(dx/\Delta x)_0$ eqn (6) can be used to obtain a plot of the refractive index across the fibre for the fixed temperature θ_0.

The accuracy of the method can be assessed by considering the errors involved in equation (6) when the specimen acts also as the standard object: (i) Using a refractometer the error in $(\mu_L)_{\theta_0}$ should be about $\pm 0{\cdot}0001$ but, as this uncertainty is common to all the measurements, it will not affect the relative refractive indices. (ii) The larger t' the smaller is the percentage error in its determination. For $t' = 0{\cdot}1$ mm this error should not exceed $\pm 2\%$ and, if an index change of 0·002 is under study, the corresponding index error is $\pm 0{\cdot}00004$. (iii) The accuracy with which $dx/\Delta x$ can be determined will depend on the fringe sharpness as well as on the absolute magnitudes of dx and Δx. It is unlikely that $dx/\Delta x$ can be found to be better than 1%; the corresponding index error is $\pm 0{\cdot}01\,\lambda/t'$, which for $t' = 0{\cdot}1$ mm is $\pm 0{\cdot}00005$. (iv) The larger the specimen thickness the smaller is the refractive index difference that can be detected (cf. equation (3)) and the more precise is the determination of the position of zero fringe displacement. With a specimen thickness of 0·1 mm the points of zero displacement can be found to within $10\,\lambda$. Taking $dx/\Delta x = 0{\cdot}01$ as the limit of detectability, the product of optical thickness uncertainty and lateral uncertainty is $(\lambda/200) \times 10\,\lambda \sim \lambda^2/20$, which is of the order of magnitude predicted from the uncertainty principle of Ingelstam (1953).

These considerations show that the relative refractive indices can be the more accurately determined the greater the specimen thickness. There is, however, a limit to this improvement. When the interferometer gap is large, the fringe sharpness will suffer a marked deterioration unless the incident light is a critically collimated beam of radiation of very narrow line width (Tolansky 1946). As a consequence of these requirements the fringe intensity will be extremely low, and for this reason the maximum specimen thickness so far dealt with has been 0·7 mm. The fringes of figure 4 were taken with an under-run high pressure mercury lamp; using 35 mm H.P.3 film an exposure time of ten minutes was required. The sharpness of the fringes under these conditions is a clear indication of the high temperature stability that was achieved.

§6. Application of the Methods

The refractive index of a specimen will often vary in a direction parallel to the microscope axis as well as in directions perpendicular thereto. Unfortunately, the interferometric methods are incapable of detecting variations of the first kind, but measure instead a mean index μ_s so defined that $\mu_s t$ represents the optical thickness of the specimen in a direction parallel to the microscope axis. It is, however, unnecessary with these methods to know the thickness t in order to determine any changes that μ_s might undergo in directions perpendicular to the microscope axis. (In polarization microscopy a compensator is used to determine the mean retardation of the specimen, i.e. mean birefringence × thickness. The mean birefringence cannot, therefore, be calculated unless the specimen thickness is known with sufficient accuracy, but this cannot be achieved in practice unless

the object has a convenient geometrical form, such as a right cylinder. Even in this instance, however, the thickness will change so rapidly near the cylinder edge that any small variation in birefringence might be completely masked.)

For certain purposes a knowledge of the mean index μ_s is inadequate and it will then be necessary to determine the refractive index at various depths of the specimen. Two approaches might then be open, the first by use of sections, the second by means of calculation from the observed values of μ_s. If the first approach is followed, serial sections will be needed, each section being submitted to a separate interferometric examination. With birefringent specimens it is the principal refractive indices that are required, and it is therefore convenient to choose the plane of the sections so as to contain two of the axes of the index ellipsoid. Each section will then produce two plane-polarized fringe systems, one for each of the two principal refractive indices. With uniaxial bodies only one set of sections is necessary, but with biaxial ones two sets, cut at right angles to one another, are needed.

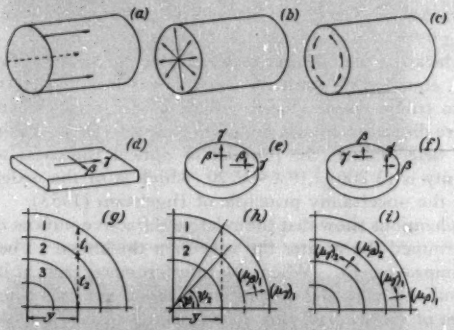


Figure 5. A right cylinder composed of uniaxial elements, whose optic axes are (*a*) axial, (*b*) radial, and (*c*) tangential. In (*d*), (*e*) and (*f*) are shown appropriate sections; μ_β and μ_γ are the principal indices of the elements, μ_β corresponding to propagation along the optic axis. The division of the cross section into zones is illustrated in (*g*), (*h*) and (*i*), where only a quarter of the circle is drawn.

Sectioning is not, however, always essential and should be avoided whenever possible because of the danger of distorting the specimen. Fortunately, objects with an axis of symmetry often offer the possibility of dispensing with sections. An example is afforded by a right cylinder composed of birefringent elements, each of which is arranged with one of its principal vibration directions parallel to the cylinder axis. When the elements are uniaxial, three interesting possibilities arise: the optic axis is (*a*) axial, (*b*) radial or (*c*) tangential (figure 5). If sections are to be used, they should be cut as shown in figures 5 (*d*), (*e*) and (*f*). If on the other hand, sections are not used, then the cylinder is placed between the optical flat with its axis at right angles to the microscope axis. It is now examined interferometrically, first with the light vibrating parallel to the cylinder axis and afterwards with it vibrating perpendicular to this axis; in this way the mean indices $(\mu_s)_{\parallel}$ and $(\mu_s)_{\perp}$ can be determined for different points across the width of the cylinder. The following analysis is then adopted.

(*a*) *Axial orientation.* The cross section is considered to be divided into p elementary annular zones (numbered 1, 2, ... p), the principal indices μ_β and μ_γ being regarded as constant within each zone (figure 5 (*g*)). If the light is vibrating parallel to the cylinder axis, the mean index $(\mu_\alpha)_{\|}$ measured at the edge of the cylinder will be equal to the index $(\mu_\gamma)_1$ of the first zone. At the distance y from the axis the optical thickness will be

$$(\mu_\alpha)_{\|}t=(\mu_\alpha)_{\|}(t_1+t_2)=(\mu_\gamma)_1t_1+(\mu_\gamma)_2t_2 \qquad \ldots\ldots(8)$$

where $(\mu_\alpha)_{\|}$ is the appropriate measured mean index and t_1 and t_2 are the appropriate calculated zone thicknesses. The index $(\mu_\gamma)_2$ can therefore be found directly. If a smaller value of y is now chosen so that three zones are involved, $(\mu_\gamma)_3$ can be calculated, and, by repetition of the procedure, one can obtain a plot of μ_γ as a function of the distance y from the axis.

If the light is vibrating at right angles to the cylinder axis, the same procedure is followed except that the indices $(\mu_\alpha)_\perp$, $(\mu_\beta)_1$, $(\mu_\beta)_2$ . . . . are now involved.

Such calculations have been applied to a cellulose fibre similar to that which gave the fringe patterns of figure 4. It was found that both μ_β and μ_γ decreased as the centre of the fibre was approached, the effective isotropic index $\mu_{iso}=\frac{1}{3}(\mu_\gamma+2\mu_\beta)$ changing from $1{\cdot}5361_3$ at the edge to $1{\cdot}5351_2$ at the centre. This change of 0·0010 represents a decrease in the optical density d of 0·2%, a possible interpretation of this result being that the inner regions of the fibre contain a smaller percentage of crystalline material. The birefringence $\mu_\gamma-\mu_\beta$ was, however, constant to within $\pm 0{\cdot}0001$, thereby showing that the degree of scatter of the molecules about the direction of orientation (i.e. about the cylinder axis) was everywhere the same.

(*b*) *Radial orientation.* For light vibrations parallel to the cylinder axis, the procedure described in (*a*) is applied to discover the changes in μ_β. For perpendicular vibrations, however, the position is more difficult, the optical thickness of the pth zone being given by the readily determinable integral

$$\int_{v_{p+1}}^{v_p} f_p\,dt$$

where

$$f_p=(\mu_\gamma)_p(\mu_\beta)_p[(\mu_\gamma)_p{}^2\sin^2\psi+(\mu_\beta)_p{}^2\cos^2\psi]^{-1/2} \qquad \ldots\ldots(9a)$$

and

$$dt=y\sec^2\psi\,d\psi, \qquad \ldots\ldots(9b)$$

ψ being the angle between the optic axis and the direction of the light vibration (figure 5 (*h*)). At the cylinder edge $(\mu_\alpha)_\perp=(\mu_\gamma)_1$. As $(\mu_\beta)_1$ has already been determined, the function f_1 can be calculated for any angle ψ. If the distance y from the axis is such that two zones are involved, then

$$(\mu_\alpha)_\perp t=\int_{v_2}^{v_1} f_1\,dt+\int_{v_1=0}^{v_1} f_2\,dt \qquad \ldots\ldots(10a)$$

where

$$t=\int_0^{v_1} dt. \qquad \ldots\ldots(10b)$$

The only unknown in eqns (10) is $(\mu_\gamma)_2$, which can therefore be calculated. Repetition of the procedure will then lead to values for $(\mu_\gamma)_3$ . . . . $(\mu_\gamma)_p$.

(*c*) *Tangential orientation.* From the results for the parallel vibrations $(\mu_\beta)_1$, $(\mu_\beta)_2$ . . . . can readily be found. The calculations for the perpendicular vibrations involve the function f'_p which is obtained from eqn (9*a*) by interchanging $(\mu_\beta)_p$

and $(\mu_\gamma)_p$. Thus, at the edge of the cylinder $(\mu_\alpha)_\perp = (\mu_\beta)_\parallel$. This quantity has already been determined and, as one has no means of finding $(\mu_\gamma)_\parallel$, the calculations break down. It is therefore necessary to work on a section as shown in figure 5 (*f*).

References

Barer, R., Ross, K. F. A., and Tkaczyk, S., 1953, *Nature, Lond.*, **171**, 721.
Brossel, J., 1946, *Nature, Lond.*, **157**, 623.
Faust, R. C., 1952, *Proc. Phys. Soc.*, B, **65**, 48.
Frey-Wyssling, A., and Wuhrmann, K., 1939, *Helv. Chim. Acta*, **22**, 981.
Hermans, P. H., 1949, *Physics and Chemistry of Cellulose Fibres* (New York : Elsevier).
Ingelstam, E., 1953, *Arkiv f. Fys.*, **7**, 309.
Tolansky, S., 1946, *Proc. Phys. Soc.*, **58**, 654.

The Dielectric Relaxation of Polar Molecules in Solution

By NORA E. HILL

Department of Physics, Bedford College, London

Communicated by H. T. Flint; MS. received 12th October 1953

Abstract. A simple theory of dielectric relaxation is proposed, based on the same model of a liquid as Andrade's theory of viscosity. It is shown that the relaxation time of a polar molecule in solution in a non-polar solvent should be related, not to the viscosity of the solvent, but to the ' mutual viscosity ' of the solute and solvent. This quantity can be calculated from the variation of the viscosity of the solution with concentration, since the viscosity η_m of a mixture containing mole fractions f_A and f_B of the two substances A and B is shown to be

$$\eta_m = f_A^2 \eta_A \frac{\sigma_A}{\sigma_m} + f_B^2 \eta_B \frac{\sigma_B}{\sigma_m} + 2 f_A f_B \eta_{AB} \frac{\sigma_{AB}}{\sigma_m}$$

where the σ's are intermolecular distances, and η_{AB}, the mutual viscosity, represents the contribution to the total viscosity of collisions between a molecule of substance A and a molecule of substance B.

The theory shows satisfactory agreement with experimental results both for pure liquids and for solutions.

§ 1. Introduction

Measurements on the electrical properties of polar liquids at microwave frequencies are usually interpreted in terms of the equation

$$\alpha = (\mu^2/3\boldsymbol{k}T)/(1+j\omega\tau) \qquad \ldots\ldots(1)$$

where α is the polarizability of the molecule due to its dipole moment μ, τ is its relaxation time, and $\boldsymbol{k}$ and T are the Boltzmann constant and absolute temperature. For an approximately spherical molecule τ has been related to the viscosity of the liquid by

$$\tau = 4\pi\eta a^3/\boldsymbol{k}T \qquad \ldots\ldots(2)$$

where a is the radius of the molecule.

This equation was first proposed by Debye (1929) for a pure polar liquid, but has since been applied by many authors to the case of polar molecules in solution in a non-polar solvent, in which case η is the viscosity of the solvent or solution. The equation was derived by assuming that a polar molecule could be regarded as a sphere rotating in a viscous medium, so that when its angular velocity is $d\theta/dt$ the molecule experiences a resistive couple

$$C = 8\pi\eta a^3 \, d\theta/dt. \qquad \ldots\ldots(3)$$

The experimental results show that this theory is fairly satisfactory for pure polar liquids, the variation of relaxation time with viscosity and temperature being of the expected form. However, if the molecular radius calculated from the molar volume is used in eqn (2) the calculated relaxation time is too high

by a factor of between five and ten (Conner and Smyth 1943, Hennelly, Heston and Smyth 1948).

When the experimental results for polar molecules in non-polar solvents are considered, the theory is less satisfactory, for the expected proportionality between the relaxation time and the solvent viscosity does not appear (e.g. Curtis, McGeer, Rathmann and Smyth 1952). The discrepancy becomes particularly marked when highly viscous solvents are used—an increase in the viscosity by a factor of about 100 may cause an increase in the relaxation time by a factor of only 2 or 3.

If one regards the process of dielectric relaxation from a microscopic point of view the origin of these discrepancies appears, for the use of the viscosity to calculate the relaxation time in the case of a solution is open to an objection which cannot be raised against Debye's calculation for a pure polar liquid. In a pure liquid every molecule is surrounded by molecules of the same kind, and the interaction between these molecules determines both the viscosity of the liquid and the relaxation time, so that, regardless of the particular argument by which a relation between the two quantities is found, it is to be expected that they will vary in a similar manner from one substance to another. But in the case of a polar solute in a non-polar solvent, the relaxation time is determined almost entirely by the interaction between solvent and solute molecules, while the viscosity of the solution is determined almost entirely by the interaction between the solvent molecules; there is therefore no reason to expect a simple relationship between the two quantities. If the solvent is a mixture of two non-polar liquids a similar argument applies, the interactions which determine the viscosity being distinct from those which determine the relaxation time.

The interaction between solvent and solute molecules which determines the relaxation time does influence the viscosity of the solution as the concentration of the polar substance is increased, and the object of the present paper is to find the connection between the variation of the viscosity of a polar–non-polar mixture with concentration and the relaxation time of the polar substance when in dilute solution in the non-polar solvent. The results have been reported briefly elsewhere (Hill 1953).

§ 2. The Viscosity of a Mixture

In order to obtain a more accurate relation between dielectric relaxation and viscous flow it is necessary to consider more closely the interactions which determine the viscous and 'inner friction' forces in a liquid, particularly in a mixture.

Theories of the viscosity of a liquid fall into two main classes: those based on a simple transfer of momentum by collision (e.g. Andrade 1934, van Wijk and Seeders 1937), and those based on the change in the intermolecular forces caused by the relative motion of the two layers (e.g. Born and Green 1947). Of the two approaches, the first lends itself more readily to the calculation of the inner friction, and the model used by Andrade has been adopted here.

2.1. *Andrade's Theory of Viscosity*

According to Andrade's theory, the molecules of a liquid can be regarded as vibrating with a certain frequency ν about equilibrium positions which vary only slowly with time. The transfer of momentum which occurs in a viscous liquid between layers moving with different velocities is due to temporary unions

(collisions) of adjacent molecules, which occur at the extremes of each swing so that for an infinitesimal time the two molecules move as one. This leads to the equation

$$\eta = \tfrac{1}{3}\nu m/\sigma \qquad \ldots\ldots(4)$$

where m is the mass of a molecule.

This equation involves the assumption that a collision takes place every time a molecule arrives at its extreme position. If we write instead

$$\eta = \tfrac{1}{3}c\nu m/\sigma \qquad \ldots\ldots(5)$$

c is the probability of such a collision and varies as $\exp(-E/kT)$, where E is the difference between the mutual potential energy of the molecules during a collision, and that when the molecules are in their mean positions.

2.2. *Extension of the Theory to a Mixture of Two Liquids*

The relaxation time of a polar molecule B in a non-polar solvent A depends on the transfer of momentum between a molecule B and a molecule A. This type of momentum transfer will also influence the viscosity of a mixture of the two liquids, therefore it is of interest to extend Andrade's theory to the viscosity of a mixture.

In this case the transfer of momentum is due to three different types of collision, which can be represented by AA, BB and AB. The first type will occur $\frac{1}{3}f_A^2 c_A 2\nu_A/\sigma_m^2$ times per second per unit area, and cause a momentum transfer $\frac{1}{2}m_A\delta v$ each time, where v is the drift velocity, thus contributing a shearing stress $\frac{1}{3}f_A^2(c_A\nu_A m_A/\sigma_m^2)\delta v$.

Similarly the second type of collision contributes a shearing stress $\frac{1}{3}f_B^2(c_B\nu_B m_B/\sigma_m^2)\delta v$.

The third type of collision occurs $\frac{1}{3}2f_A f_B c_{AB}(\nu_A+\nu_B)/\sigma_m^2$ times per second per unit area, and causes a momentum transfer $\{m_A m_B/(m_A+m_B)\}\delta v$ each time, thus contributing a shearing stress

$$\tfrac{1}{3}2f_A f_B c_{AB}\frac{m_A m_B}{m_A+m_B}\frac{\nu_A+\nu_B}{\sigma_m^2}\delta v.$$

The total shearing stress is

$$\left\{f_A^2\frac{c_A\nu_A m_A}{3\sigma_m^2}+f_B^2\frac{c_B\nu_B m_B}{3\sigma_m^2}+2f_A f_B\frac{c_{AB}(\nu_A+\nu_B)m_A m_B}{3\sigma_m^2(m_A+m_B)m}\right\}\delta v.$$

Hence the viscosity is

$$\eta_m = f_A^2\eta_A\sigma_A/\sigma_m + f_B^2\eta_B\sigma_B/\sigma_m + 2f_A f_B\eta_{AB}\sigma_{AB}/\sigma_m \qquad \ldots\ldots(6)$$

where η_A and η_B are the viscosities of the pure liquids A and B and η_{AB} the 'mutual viscosity' is defined by

$$\eta_{AB} = \tfrac{1}{3}c_{AB}\frac{(\nu_A+\nu_B)}{\sigma_{AB}}\frac{m_A m_B}{m_A+m_B}, \qquad \ldots\ldots(7)$$

σ_{AB} being the mean distance between a molecule of type A and an adjacent molecule of type B in the mixture.

2.3. *Comparison with Experimental Data*

A good summary of the types of viscosity–concentration curve which are found experimentally, and the relations which have been suggested to account for them is given by Jaeger (1938).

The experimental curves fall mostly into three classes: (i) Those which sag below the straight line $\eta_m = f_A\eta_A + f_B\eta_B$; this is the most common type. (ii) Those which rise above the straight line; η_m may become greater than both η_A and η_B. (iii) Those which exhibit a point of inflection; these are relatively

uncommon, and in many of the reported cases (e.g. the ether–chloroform mixture quoted by Jaeger) the point of inflection is present because the viscosity is plotted against the weight or volume concentration rather than against the molar concentration.

A large number of empirical or semi-empirical relations for the viscosity of a mixture have been proposed, none of which is universally successful and many of which are incapable of predicting cases (ii) and (iii) above. The number of relations which have a theoretical basis is much smaller. Stuart (1948) proposed a formula derived from van Wijk and Seeders' theory of viscosity. In the notation used above this is

$$\eta_m = w_A^2 \eta_A \rho_m^2/\rho_A^2 + w_B^2 \eta_B \rho_m^2/\rho_B^2 + 2w_A w_B \eta_{AB} \quad \ldots\ldots(8)$$

where w_A and w_B are the weight concentrations of the two components, and the ρ's are densities. On converting to mole concentrations this becomes

$$\eta_m = f_A^2 \eta_A (\sigma_A/\sigma_m)^6 + f_B^2 \eta_B (\sigma_B/\sigma_m)^6 + 2f_A f_B (\sigma_A/\sigma_m)^3 (\sigma_B/\sigma_m)^3 (\rho_A \rho_B/\rho_m^3)\eta_{AB}. \quad \ldots\ldots(9)$$

The η_{AB} of Stuart's relation is not, of course, defined in the same way as the η_{AB} in the theory above, but it has a similar significance.

Yang (1949), applying the theory of simple liquids developed by Born and Green to the case of a mixture, also derives an expression containing three terms, the third term taking into account the interaction between molecules of different kinds. Yang abandons the task of numerical calculation from the formula, but since Born and Green's equation for the viscosity of a pure liquid reduces to an equation very similar to Andrade's if certain assumptions are made, it seems probable that in similar circumstances Yang's equation will tend to eqn (6) above. Balázs (1951), using as a model structureless particles moving in a periodic field, obtained

$$\ln \eta_m = f_A \ln \eta_A + f_B \ln \eta_B - \beta f_A f_B w_{AB}/\boldsymbol{k}T \quad \ldots\ldots(10)$$

where w_{AB} is the energy of mixing and β is a fraction which is the same for all liquids.

In table 1 these three relations ((*a*) Hill, (*b*) Stuart, (*c*) Balázs)

$$\eta_m = f_A^2 \eta_A \sigma_A/\sigma_m + f_B^2 \eta_B \sigma_B/\sigma_m + 2f_A f_B \eta_{AB} \sigma_{AB}/\sigma_m \quad (a)$$

$$\eta_m = f_A^2 \eta_A (\sigma_A/\sigma_m)^6 + f_B^2 \eta_B (\sigma_B/\sigma_m)^6 + 2f_A f_B (\rho_A \rho_B/\rho_m^2)(\sigma_A/\sigma_m)^3 (\sigma_B/\sigma_m)^3 \eta_{AB} \quad (b)$$

$$\ln \eta_m = f_A \ln \eta_A + f_B \ln \eta_B - \beta f_A f_B w_{AB}/\boldsymbol{k}T \quad (c)$$

are compared with the experimental results for various pairs of substances, taken from the *International Critical Tables*. Columns 2, 3, and 4 show the adjustable

Table 1

Mixture	Parameter			Error (%)		
	(*a*)	(*b*)	(*c*)	(*a*)	(*b*)	(*c*)
$C_6H_6 + C_6H_5Br$	4·67	0·623	−0·0823	2·8	6·9	1·3
$C_6H_6 + C_6H_5Cl$	4·05	0·651	−0·0507	1·1	2·3	0·9
$CH_3NO + CH_3COOH$	26·6	6·28	−1·255	1·4	3·0	6·4
$CS_2 + C_6H_5OH$	1·75	(−0·42)	−0·090	4·5	21·7	2·0
$HCOOH + CH_3NO$	14·4	3·64	−0·521	1·3	1·3	3·1
$CHCl_3 + C_6H_5Br$	5·76	1·045	−0·0756	1·8	3·9	1·1
$CCl_4 + C_2H_6O$	3·03	0·938	−0·036	2·5	3·9	1·8

The parameters are :

(*a*) Hill, $\eta_{AB}\sigma_{AB} \times 10^{16}$ poise cm; (*b*) Stuart, $\eta_{AB}\rho_A\rho_B/\rho_m^2$ cP; (*c*) Balázs, (log e) $\beta w_{AB}/\boldsymbol{k}T$.

parameters calculated in each case from the viscosity of the mixture at a molar concentration of about 50%. Columns 5, 6 and 7, show the mean deviation (calculated without regard to sign) of the calculated values of the viscosity from the experimental values. The overall mean and maximum errors are:

Hill	mean 2·2%,	maximum 9·3%.
Stuart	mean 6·1%,	maximum 63·0%.
Balázs	mean 2·4%,	maximum 11·5%.

The relationships are shown graphically for formamide–acetic acid and carbon disulphide–ethyl alcohol in figure 1 (*a*) and (*b*). Stuart's relationship evidently

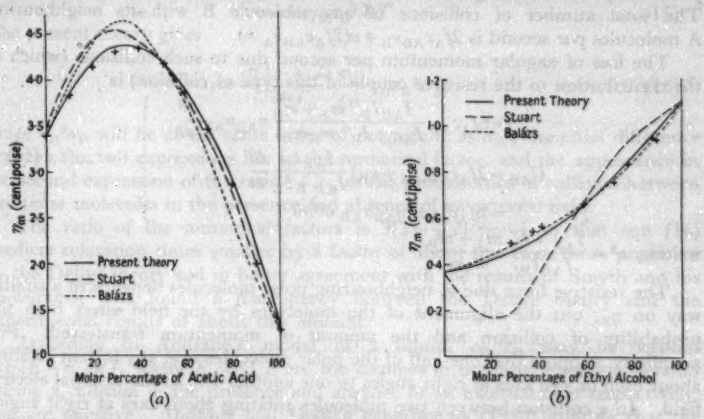


Figure 1. The three theoretical viscosity–concentration curves, and the experimental values.
(*a*) Formamide and acetic acid. (*b*) Carbon disulphide and ethyl alcohol.

has too great a dependence on the intermolecular distances, for in the case of the carbon disulphide–ethyl alcohol mixture, where the variation of this quantity is large (4·01 to 5·40Å), η_{AB} has to be given a negative value, which has no physical significance, and even then there is a large deviation from the experimental values. Balázs' theory and the present one give very similar results, except in the case of the formamide–acetic acid mixture, where the viscosity concentration curve is very concave to the concentration axis, in which case the fit of Balázs' relation is not so good.

On the whole it appears that relation (*a*) is in sufficiently good agreement with experiment to justify the use of the mutual viscosity calculated from it as a measure of the interaction between solvent and solute molecules.

§3. The Mutual Viscosity and the Inner Friction

The resistive force which a rotating polar molecule experiences in a non-polar liquid depends chiefly on the mutual viscosity of the two liquids, if the resistive force arises in a similar way to the viscous forces. The polar molecule has the probability c of a collision every time it or one of its neighbours is at the extreme

end of its swing. Suppose the polar molecule B (assumed rigid) is rotating with angular velocity $d\theta/dt$ and collides with a neighbouring non-polar molecule A. Since, on the average, the non-polar molecules have no angular momentum, conservation of momentum requires that the polar molecule loses momentum $\{I_{AB}I_B/(I_{AB}+I_B)\}\, d\theta/dt$, where I_B is the moment of inertia of the polar molecule about its centre of mass and I_{AB} is the moment of inertia of the molecule A about the centre of mass of B at the instant of collision.

Such a momentum loss occurs $2f_Ac_{AB}\nu_B$ times per second due to the oscillations of B, and $2f_Ac_{AB}\nu_A/n$ times per second due to the oscillations of A, where n is the number of near neighbours (because any molecule has the probability c of colliding with *one* of its neighbours at every extreme position). The total number of collisions of any molecule B with its neighbouring A molecules per second is $2f_Ac_{AB}\nu_B+n(2f_Ac_{AB}\nu_A/n)$.

The loss of angular momentum per second due to such collisions (which is the contribution to the resistive couple of this type of collision) is

$$\left.\begin{aligned} C_{AB} &= 2f_Ac_{AB}(\nu_A+\nu_B)\frac{I_{AB}I_B}{I_{AB}+I_B}\frac{d\theta}{dt} \\ &= 6f_Ac_{AB}\kappa^2\eta_{AB}\sigma_{AB}\, d\theta/dt \end{aligned}\right\} \quad \text{.....(11)}$$

writing $\kappa^2=\dfrac{I_{AB}+I_B}{I_{AB}I_B}\dfrac{m_A+m_B}{m_Am_B}$.

The resistive force due to neighbouring polar molecules depends in a similar way on η_B, but the alignment of the molecules by the field alters both the probability of collision and the amount of momentum transferred. For simplicity, suppose that one-half of the polar molecules is at any instant rotating about each of two axes at right angles to one another and to the external electric field. In a collision between two molecules rotating about axes at right angles to one another, both molecules lose angular momentum $I_B(1-1/\sqrt{2})\, d\theta/dt$. This is the case for one-half the total number of collisions. For the remaining collisions the momentum loss is 0 or $I_B\, d\theta/dt$ according as the rotations are in the same or opposite senses. The mean momentum loss per collision due to collisions between polar molecules is $\frac{1}{4}\{2I_B(1-1/\sqrt{2})+I_B\}\, d\theta/dt$. The number of collisions per second between polar molecules is $4f_B\nu_Bc_B'$, where c_B' is the probability of collision, and may vary with the degree of alignment of the molecules, and therefore with the frequency of the external field. Some recent results of Andrade and Dodd (1950) on the effect of an electric field on the viscosity of a polar liquid indicate that c_B' does not differ much from c_B, the probability in the absence of a field.

The resistive couple due to the polar molecules is

$$\begin{aligned} C_B &= f_B\nu_Bc_B'I_B(3-\sqrt{2})\, d\theta/dt \\ &= 3(3-\sqrt{2})f_B(c_B'/c_B)\kappa_B^2\eta_B\sigma_B\, d\theta/dt. \end{aligned} \quad \text{.....(12)}$$

Therefore the total resistive couple is

$$C=\{f_A6\kappa^2\eta_{AB}\sigma_{AB}+f_B3(3-\sqrt{2})(c_B'/c_B)\kappa_B^2\eta_B\sigma_B\}\, d\theta/dt \quad \text{.....(13)}$$

and the relaxation time is

$$\tau=\frac{1}{2kT}\{f_A6\kappa^2\eta_{AB}\sigma_{AB}+f_B3(3-\sqrt{2})(c_B'/c_B)\kappa_B^2\eta_B\sigma_B\}. \quad \text{.....(14)}$$

Except in the case of spherically symmetrical molecules, I_A and I_{AB} have a range of possible values, so that τ is not sharply defined. This will cause a broadening of the dispersion curve or, if the moments of inertia about different axes differ widely, the appearance of more than one region of dispersion in the frequency spectrum.

3.1. *Comparison with the Experimental Results*

(*a*) *For pure polar liquids.*

For the relaxation time of a pure polar liquid Debye gave

$$\tau = 4\pi\eta_B a_B^3/kT. \qquad \ldots\ldots(15)$$

The present theory gives

$$\tau = \frac{3(3-\sqrt{2})}{2kT}\frac{c_B'}{c_B}\eta_B\kappa_B^2\sigma_B. \qquad \ldots\ldots(16)$$

Since $\kappa_B^2\sigma_B$ will be of the same order of magnitude as a_B^3, the chief difference between the two expressions lies in the numerical factor, and the appearance in the second expression of the ratio c_B'/c_B of the probabilities of collision between two polar molecules in the presence and absence of an external field.

The ratio of the numerical factors is $3(3-\sqrt{2})/8\pi \sim \frac{1}{5}$ so that eqn (16) predicts relaxation times smaller by a factor of about five than those predicted by the Debye theory and in better agreement with the results of Smyth and his co-workers, who found a discrepancy between the Debye theory and the experimental results of about this amount.

According to Andrade's model of a liquid, c_B'/c_B may be written $\exp(E-E')/kT$, where E' and E are the changes in potential energy occurring during a collision in the presence and absence of an external field respectively, and are negative. Andrade and Dodd's experiments (1950) indicate that the deviation of c_B'/c_B from unity is positive, and is certainly within the error of relaxation time measurements. The experimental evidence on this point is inconclusive. Whiffen and Thompson (1946) found $|E|$ slightly greater than $|E'|$ for toluene, p-cymene and o-xylene, although the difference was almost within the experimental error, while Hennelly, Heston and Smyth (1948) found E'/E about 0·6 for a series of organic halides. The latter set of results, however, refers chiefly to large molecules which do not satisfy the assumptions of the theory.

Equation (16) agrees with the Debye equation in predicting an increase of τ with increasing viscosity and molecular size.

(*b*) *For a polar solute in a non-polar solvent.*

For a very dilute solution of a polar substance in a non-polar solvent the Debye relation is

$$\tau = 4\pi\eta_A a_B^3/kT \qquad \ldots\ldots(17)$$

and the present theory gives

$$\tau = \frac{3}{kT}\frac{I_{AB}I_B}{I_{AB}+I_B}\frac{m_A+m_B}{m_A m_B}\eta_{AB}\sigma_{AB}. \qquad \ldots\ldots(18)$$

The chief difference here is that the viscosity of the solvent in the Debye relation is replaced by the mutual viscosity of the solute and solvent in eqn (18). This offers an immediate explanation of the different relaxation times found for one solute in a variety of solvents having the same viscosity, and for the relatively low relaxation times found for solutes in high viscosity solvents.

In table 2 values of η_{AB} calculated from results in the *International Critical Tables* for solutes in benzene and from measurements made by the writer in other cases are compared with the relaxation times measured by Whiffen and Thompson (1946), Whiffen (1950) and Curtis, McGeer, Rathmann and Smyth (1952) for various pairs of substances. The ratios τ/η_{AB} and τ/η are shown for comparison. The last column shows l, the mean distance of closest approach of the molecules,

Table 2

Solute	Solvent	$\tau\times10^{12}$ (sec)	η_{AB} (cP)	$(\tau/\eta_{AB})\times10^{10}$	$(\tau/\eta)\times10^{10}$	l(Å)
Chlorobenzene	Benzene	7·8	0·73	10·7	12·8	0·94
	Paraffin	40	3·50	11·4	0·2	—
Bromobenzene	Benzene	10·6	0·90	11·8	16·3	1·36
Nitrobenzene	Benzene	12·8	0·82	15·5	19·7	1·75
	Cyclo-hexane	9·5	1·02	9·3	9·8	0·25
	Carbon tetrachloride	15·3	1·19	12·8	15·8	0·87
Ethyl benzoate	Benzene	12·9	0·96	13·4	19·9	—
Acetone	Benzene	3·2	0·39	8·2	4·9	—
Chloroform	Benzene	7·1	0·61	11·7	10·9	neg.
	Cyclo-hexane*	3·2	0·31	10·3	3·3	1·0
	Carbon tetrachloride*	5·0	0·59	8·5	5·2	0·9

* For these mixtures the value of η_{AB} calculated from the mixture equation varies slightly with concentration. The value used in the table is the extrapolated value for zero concentration of the polar substance.

calculated from τ/η_{AB} and the estimated radii of gyration of the molecules. Considering the simplicity of the model, and the extreme sensitivity of l to the numerical factor in the equation for τ, the values of l are very reasonable, except for chloroform in benzene, for which the estimated moments of inertia are such that l becomes negative. In the case of medicinal paraffin $\frac{1}{3}kT\tau/\eta_{AB}\sigma_{AB}$ tends to I_B^2, owing to the large mass of the paraffin molecule. The radius of gyration of the chlorobenzene molecule calculated in this way is 1·5 Å, compared with 1·85 Å calculated from the structure of the molecule. There is some doubt about the value to be assigned to the mutual viscosity in this case, as the mixture law is not followed, probably because of the size and lack of rigidity of the paraffin molecules. The mutual viscosity was calculated from the variation of viscosity at low concentration of the paraffin since it is here that the conditions approach most closely to those required by the theory. Considering these approximations no great significance can be attached to the agreement between the two values of the radius of gyration of the chlorobenzene molecules, but it does show that the present theory is capable of giving results of the correct order of magnitude, in a case where the Debye relation fails completely.

The results are displayed graphically in figures 2 and 3 which show τ plotted as a function of η_A and of η_{AB} respectively. From these graphs it can be seen

that τ is not proportional to η_A as required by the Debye formula; but the graph of τ against η_{AB} does approximate to a straight line as required by eqn (18).

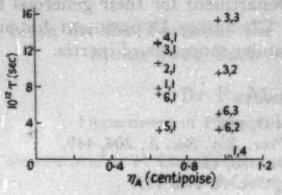


Figure 2.

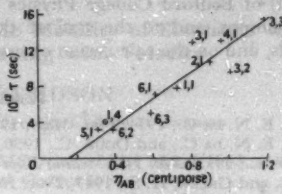


Figure 3.

Figure 2. Relaxation time and solvent viscosity for various polar–non-polar solutions. The numbers indicate solute, solvent:

Solutes: 1, C_6H_5Cl; 2, C_6H_5Br; 3, $C_6H_5NO_2$; 4, $C_2H_5COOC_2H_5$; 5, $CH_3CO.CH_3$; 6, $CHCl_3$.
Solvents: 1, C_6H_6; 2, C_6H_{12}; 3, CCl_4; 4, Medicinal paraffin.
For the point marked by a circle, both scales should be multiplied by 200.

Figure 3. Relaxation time and mutual viscosity. The solutions are numbered in the same way as in figure 2. For the point marked by a circle both scales should be multiplied by 10.

§ 4. Conclusion

The relationship for the relaxation time, derived from Andrade's model of a liquid, shows better agreement with experiment than the Debye relation, both for pure liquids and for solutions. The chief point of the theory is that the relaxation time of a polar substance in solution should be related, not to the viscosity of the solvent, but to the 'mutual viscosity' of the solute and solvent, and this is not dependent on the particular model used; a similar relationship will arise from any theory in which the viscosity is pictured as being due to intermolecular action of any sort, rather than to the transfer of momentum by the changing positions of the molecules themselves. It seems probable that in a liquid there will be a small contribution to the viscosity due to the diffusion of molecules from the faster to the slower moving parts of the liquids and vice versa. Such a mechanism would make no contribution to the inner friction and may account for the tendency of the graph of τ against η_{AB} to cut the viscosity axis to the right of the origin. A similar effect occurs in the values of τ deduced by Saxton (1952) from the results of Lane and Saxton (1952) for water, methyl and ethyl alcohol over the temperature range $-10°$C to 50°C. If graphs are plotted of τ against η, the alcohols give straight lines which intersect the viscosity axis at 2·8 millipoise for ethyl and 2·4 millipoise for methyl alcohol. Water gives a slight curve, concave towards the viscosity axis and intersecting it between 0 and 1 millipoise. A slight curvature is to be expected, as the diffusive viscosity term will increase with increasing temperature, and tend to zero at the freezing point.

ACKNOWLEDGMENTS

The author wishes to express her thanks to Professor H. T. Flint, under whose direction this work was carried out, to the staff (both academic and technical) of Bedford College Physics Department for their generous help and encouragement, and to the staff of the Chemistry Department for advice on chemicals, and on the calculation of molecular moments of inertia.

REFERENCES

ANDRADE, E. N. DA C., 1934, *Phil. Mag.*, **17,** 497, 698.
ANDRADE, E. N. DA C., and DODD, C., 1950, *Proc. Roy. Soc.* A, **204,** 449.
BALÁZS, N. L., 1951, *Rec. Trav. Chim. Pays-Bas*, **70,** 412.
BORN, M., and GREEN, H. S., 1947, *Proc. Roy. Soc.* A, **190,** 455.
CURTIS, A. J., MCGEER, P. L., RATHMANN, G. B., and SMYTH, C. P., 1952, *J. Amer. Chem. Soc.*, **74,** 644.
CONNER, W. P., and SMYTH, C. P., 1943, *J. Amer. Chem .Soc.*, **65,** 382.
DEBYE, P., 1929, *Polar Molecules* (New York : Chemical Catalog Co.).
HENNELLY, E. J., HESTON, W. M., Jr., and SMYTH, C. P., 1948, *J. Amer. Chem. Soc.*, **70,** 4102.
HILL, N. E., 1953, *Nature, Lond.*, **171,** 836.
JAEGER, F. M., 1938, *Second Report on Viscosity and Plasticity* (Amsterdam : Nordemann).
LANE, J. A., and SAXTON, J. A., 1952, *Proc. Roy. Soc.* A, **213,** 400.
SAXTON, J. A., 1952, *Proc. Roy. Soc.* A, **213,** 473.
STUART, H. A., 1948, *Z. Naturforsch.*, **3a,** 196.
VAN WIJK, W. R., and SEEDERS, W. A., 1937, *Physica*, **4,** 1073.
WHIFFEN, D. H., 1950, *Trans. Faraday Soc.*, **46,** 130.
WHIFFEN, D. H., and THOMPSON, H. W., 1946, *Trans. Faraday Soc.*, **42A,** *Discussion on Dielectrics*, p. 114.
YANG, L. M., 1949, *Proc. Roy. Soc.* A, **198,** 471.

RESEARCH NOTES

A Note on the Formula for the Mobility of Electrons with Mean Free Path varying with Velocity

By P. M. DAVIDSON
Department of Physics, University College of Swansea

Communicated by F. Llewellyn Jones; MS. received 18*th March* 1953, *and in amended form* 5*th October* 1953

§ 1. Introduction

In many discussions of the properties of ionized gases a formula for the electronic drift velocity $\overline{W}$ is required, and on the basis of ideal gas kinetic theory the formula commonly employed is (see, for example, Townsend 1936)

$$\overline{W}e = \tfrac{4}{3}(E/m)\lambda\overline{c^{-1}} \qquad \ldots\ldots(1)$$

where e is the charge on the electron, m its mass, E the electric field, and λ is the mean free path, assumed to be independent of the velocity c; a symbol $\overline{\phi}$ means $(\int \phi\, dn)/n$, the integral being taken over the particular velocity distribution present. This formula has been used, for example, to estimate $\lambda\overline{c^{-1}}$ from the observed $\overline{W}$. In many gases, however, the Townsend–Ramsauer effect cannot be neglected, and it is necessary to take account of the variation of λ with c. It has often been maintained that for this purpose the factor $\lambda\overline{c^{-1}}$ in (1) has merely to be replaced by $\overline{\lambda c^{-1}}$. It will be shown that this rests on an erroneous argument, and a correct generalization will be derived below.

§ 2. The Generalized Formula

The correct generalization of (1) may be obtained very easily from a well-known theorem, of which a brief but adequate derivation will first be given, since it is little longer than a definition of the symbols. Consider n electrons, each of charge e and mass m, diffusing in a uniform field of magnitude E, of great extent; thus when their density $\rho(c, \theta)$ in a velocity diagram has become constant in time we have (see, for example, Cahn 1949)

$$\frac{c\rho}{\lambda} - R = -\gamma\left(\cos\theta\frac{\partial\rho}{\partial c} - \frac{\sin\theta}{c}\frac{\partial\rho}{\partial\theta}\right) \qquad \ldots\ldots(2)$$

where γ is eE/m, and R is the rate of increase of $\rho(c, \theta)$ due to collisional transference from other regions of the diagram; the polar axis has the direction of the electric field. It is convenient to write $\rho(c, \theta)$ in the form $\rho_0(c) + \rho_1(c, \theta)$ where ρ_1 is zero when averaged over a spherical shell of radius c. If we write W for the mean axial velocity of the electrons with velocity c (strictly speaking, between c and $c + dc$), we have $W = di/dn$ where dn is $4\pi c^2\rho_0 dc$ and di is $2\pi c^3 dc\rho_1 \sin\theta\cos\theta\, d\theta$ integrated with respect to θ. We will at first confine ourselves to cases in which the collisional processes are of a kind which gives R spherical symmetry and in which ρ_1 is small compared with ρ_0 (and thus W, which is di/dn small compared with c), except perhaps in extreme parts of the

velocity range, of no importance for our purpose. Multiplying (2) by $2\pi\lambda c^2 dc \sin\theta \cos\theta\, d\theta$ and integrating with respect to θ we obtain

$$di = -\frac{4\pi\gamma}{3}\lambda c^2 \frac{d\rho_0}{dc}\,dc \qquad \ldots\ldots(3)$$

omitting from the right-hand side its smaller terms. Thus, dividing by dn,

$$W = -\tfrac{1}{3}\gamma\lambda\{d\rho_0/dc\}/\rho_0. \qquad \ldots\ldots(4)$$

Since W is di/dn, we see that i/n is $\overline{W}$, and is thus obtained by putting a bar over (4).

These, as remarked above, are quite well-known results. To get from them an expression for comparison with (1) we have merely to integrate (3) by parts; we then have

$$\frac{i}{n} = \frac{\gamma}{3n}\int 4\pi\rho_0 \frac{d(\lambda c^2)}{dc}\,dc = \frac{\gamma}{3n}\int \frac{d(\lambda c^2)/dc}{c^2}\,dn = \frac{\gamma}{3}\overline{\left\{\frac{d(\lambda c^2)/dc}{c^2}\right\}}$$

that is,

$$\overline{W} = \frac{eE}{3m}\overline{\left\{\frac{2\lambda}{c} + \frac{d\lambda}{dc}\right\}}. \qquad \ldots\ldots(5)$$

A simple example of a varying λ is that in which λ/c is a constant, say T. Inserting in (5) we have $\overline{W} = eET/m$. Thus if we had used the erroneous generalization of (1) we should be wrong by a factor of $\frac{2}{3}$ in this case.

Equation (5) shows the correct generalization of (1) and disproves the incorrect generalization, but it is of no other interest, for if we have enough data to evaluate $\overline{W}$ from (5) we can as easily evaluate it by merely putting a bar over (4). Thus it would hardly be surprising if (5) did not appear in the literature; actually, however, it will be found in the footnotes to a paper by Smit (1936) (in which the subject of mobility plays only an incidental part), but it is given without proof, and is accompanied by the erroneous statement that the right-hand side without the bar is the drift velocity of the electrons of velocity c. That drift velocity is given, as we saw, by (4). We obtained (5) by partial integration: its right-hand side without the bar differs from (4) by $\gamma\{d(\lambda c^2\rho_0)/dc\}/3c^2\rho_0$ and it is only when averaged that they are equal.

§ 3. Physical Interpretation

It is of interest to consider the true physical significance of the expression on the right-hand side of (5) without the bar. Suppose we had defined W as the existing mean axial velocity, not of the electrons in the range c to $c+dc$, but of the electrons which were in that range immediately after their last collision. If the numbers in the two groups are $f(c)\,dc$ and $g(c)\,dc$, the functions $f(c)$ and $g(c)$ only differ slightly; but though the number of electrons in the two groups is practically the same, they are not the same electrons, and (as may be seen on visualizing the motions) they give quite different contributions to the current. $\overline{W}$ is still i/n, but now (4) is not valid without bars, whereas (5) is (except in a range of small c, of no importance). This may be proved in several ways, for example, by using the velocity diagram equation, analogous to (2), for the electrons whose velocities were in a given infinitesimal range immediately after their last collision. In this diagram ρ_1 will not be small compared with ρ_0 but, owing to the very narrow velocity distribution, an integration throughout velocity space can readily be made, showing that W for this group of particles

is the right-hand side of (5) without the bar. A longer, but perhaps more easily visualized, proof can be obtained by modifying a calculation well known in the case of λ constant. In this calculation we consider the various paths traversed by the n electrons in a given long time. Of these paths, all those whose initial velocity is c_0 (or strictly speaking in an infinitesimal range around it) are drawn, radiating from a point P. To calculate the total displacement of the electrons due to their having traversed these slightly curved paths, whose directional distribution at P is uniform, and whose lengths are, of course, distributed according to the law $dN/ds = -N/\lambda(c)$, they are compared with a set of straight lines, which will be called 'unperturbed paths', of equal number, traversed with uniform velocity c_0, with uniform directional distribution, and with lengths distributed according to a law $dN/ds = -N/\lambda(c_0)$. If λ is constant there is evidently a one-to-one correspondence between perturbed and unperturbed paths having the same direction at P and the same s. But this will not be so if λ varies with c.* For example, if λ/c is a constant T, both laws reduce to $dN/dt = N/T$, where t is the duration of the path, and there is thus a one-to-one correspondence between perturbed and unperturbed paths having the same direction at P and the same duration. When λ is an arbitrary function of c nothing is gained by introducing the unperturbed set; the calculation, without it, is quite straightforward.

In conclusion, it may be observed that if the assumption that R has spherical symmetry is abandoned it is well known that (3) will, provided certain conditions are satisfied, remain valid if the λ in it is replaced by the mean free path for momentum transfer (see, for example, Cahn 1949 and references there given). The same is therefore true of (5), but it will be found that in this more general case the right-hand side of (5) without the bar no longer has a simple physical interpretation as a drift velocity.

References

Cahn, J. H., 1949, *Phys. Rev.*, **75**, 293.
Smit, J. A., 1936, *Physica*, **3**, 543.
Townsend, J. S. E., 1936, *Phil. Mag.*, **22**, 145.

Velocity of Discharge Propagation in Self-Quenching Geiger–Müller Counters

By P. A. C. Mortier and J. F. Roose

Laboratorium voor Natuurkunde, Rijksuniversiteit, Ghent, Belgium

MS. received 7th October 1953

We have measured the velocity of discharge propagation v in a counter of 135 cm length with divided cathodes, sketched in figure 1. The cathodes are brass tubes, 1 mm thick, internal diameter 22 mm. The anode is a central wire in molybdenum, diameter 0·08 mm. The cathodes are carefully aligned and held at a mutual distance of about 0·2 mm by means of Pyrex tubes sealed on to the brass with araldite. Before mounting, the counter

* This fact is always overlooked in derivations of the erroneous formula $\overline{W} = \frac{2}{3}(eE/m)\overline{\lambda c^{-1}}$.

was rinsed with carbon tetrachloride, and evacuated at 10^{-5} cm Hg for several days. For obvious reasons it could not be baked.

The self-quenching fillings used consist of spectroscopically pure noble gases and ethyl alcohol vapour. The alcohol was distilled into its permanent reservoir under vacuum. The pressure was measured with a mercury manometer.

Discharges were initiated in A by a collimated beam of γ-rays, proceeding from a weak Ra source. At tensions above starting voltage these discharges travel down the whole counter. The pulse in section A is used to trigger a single sweep time base connected to the X-deflection plates of an oscilloscope, to the Y-deflection plates of which are applied the pulses generated in sections B and C. The distance between the leading edges of these pulses measures the time interval required by the discharge to travel over a known length of the counter, and hence its propagation velocity. The purpose of section D is to introduce a time delay before the application of the pulse in B to the Y plates. This time delay is useful when the initial part of this pulse has to be displayed on the screen. The pulses were recorded photographically.

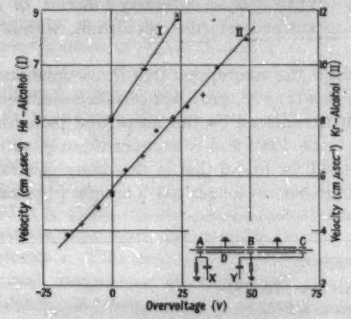


Figure 1. Velocity of discharge propagation in He (9 cm Hg)–alcohol (1 cm Hg) and Kr (9 cm Hg)–alcohol (1 cm Hg) mixtures.

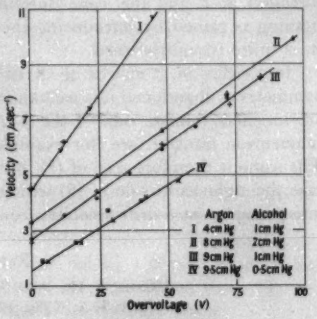


Figure 2. Velocity of discharge propagation in various A–alcohol mixtures.

The time base is a slightly modified version of the circuit of McMullan (1946). It was calibrated with a quartz-stabilized oscillator and is linear to within 3% over the range of speeds encountered.

The starting voltage V_s is defined as the voltage at which practically all pulses initiated in A generate a discharge spreading through the whole counter. It was reproducible to within 2 volts for one and the same filling, and to within 5 volts for independent fillings. It is of course possible to use either one of the several sections as a separate counter by adjusting the cathode potentials of the remaining ones so as to put them out of action. Again the whole tube may be used as one single counter. When operating in this way it is possible to define and measure V_s conventionally as the voltage where all pulses start to have the same amplitude. Starting voltages measured by this procedure coincided with the one defined above, both for the separate sections of the tube and for the tube used as one long counter.

The plateau of the counter was very narrow for the Ne–alcohol and the Xe–alcohol mixtures. For the other fillings it had a length up to about 100 volts in some cases.

In figures 1 and 2 the velocity of discharge propagation v is plotted against the overvoltage (working voltage minus starting voltage) for one He–alcohol, one Kr–alcohol, and four A–alcohol mixtures. In one case some measurements are included under starting voltage (figure 1). Different indications refer to independent measurements, made in every case at least 24 hours after filling to allow for thorough mixing. The recorded values are arithmetic means of about thirty measurements. The individual values spread somewhat more than can be accounted for by the lack of perfect linearity of the time base. Similar fluctuations are reported by Saltzmann and Montgomery (1950) and are probably significant for the mechanism of discharge propagation. The starting voltages and the velocities at starting voltage are recorded in tables 1 and 2. The results for the A–alcohol and He–alcohol mixtures are in good agreement with earlier published values (Hill and Dunworth 1946, Laufer 1950, Balakrishnan and Craggs 1950). For the remaining fillings we have not found references in the literature.

Table 1. Velocity of Discharge Propagation at Starting Voltage in Mixtures of Noble Gas (9 cm Hg) and Ethyl Alcohol (1 cm Hg)

	He	Ne	A	Kr	Xe
V_s (v)	93_8	87_8	84_0	103_4	115_7
v (cm μsec^{-1})	$5{\cdot}05 \pm 0{\cdot}05$	$5{\cdot}5 \pm 0{\cdot}5$	$2{\cdot}65 \pm 0{\cdot}10$	$5{\cdot}3 \pm 0{\cdot}1$	$6{\cdot}6 \pm 0{\cdot}2$

Table 2. Velocity of Discharge Propagation at Starting Voltage in Mixtures of Argon and Ethyl Alcohol

Argon (cm Hg)	9·5	9	8	4
Alcohol (cm Hg)	0·5	1	2	1
V_s (v)	71_6	84_0	95_2	82_9
v (cm μsec^{-1})	$1{\cdot}6 \pm 0{\cdot}1$	$2{\cdot}65 \pm 0{\cdot}10$	$3{\cdot}3 \pm 0{\cdot}1$	$4{\cdot}5 \pm 0{\cdot}1$

ACKNOWLEDGMENT

We wish to thank Prof. Dr. R. Moens for his interest in this investigation.

REFERENCES

BALAKRISHNAN, C., and CRAGGS, J. D., 1950, *Proc. Phys. Soc.* A, **63**, 358.
HILL, J. M., and DUNWORTH, J. V., 1946, *Nature, Lond.*, **158**, 833.
LAUFER, A. R., 1950, *Rev. Sci. Instrum.*, **21**, 244.
MCMULLAN, D., 1946, *Electron. Engng*, **28**, 21.
SALTZMANN, H., and MONTGOMERY, C. G., 1950, *Rev. Sci. Instrum.*, **21**, 548.

Scattering of Light in Photographic Emulsion Layers*

By G. C. FARNELL
Kodak Research Laboratories, Wealdstone, Harrow, Middx.

MS. received 8th *May* 1953

In the course of a comparative examination of pure silver bromide and silver bromo-iodide photographic emulsion layers measurements were made of the decrease of intensity of various monochromatic radiations through such layers. The method employed was to expose two pieces of film in optical contact, emulsion to emulsion, separate them, and develop them together. On plotting the characteristic curves of the two layers exposed in this way their separation parallel to the log-exposure axis provides a fairly accurate measure of the logarithm of the ratio of intensity of radiation at the top and bottom of a single layer (Webb 1948), a quantity which can be conveniently termed the unexposed emulsion density Δ.

It was considered to be of interest to use the value of Δ and the mass of silver bromide per unit area of the silver bromide emulsions for the deduction of an absorption coefficient of the halide in emulsion form, and to compare the values so obtained with those deduced by Slade and Toy (1920) for crystalline sheets of silver bromide. The results of such comparisons were quite surprising. The absorption coefficients deduced from Δ-values of several emulsions to 4360 Å radiation were 12–20 times greater than the value obtained by Slade and Toy.

Measurements of Δ for a number of wavelengths immediately showed that this ratio was wavelength dependent. A typical set of results is presented in table 1.

Table 1. Comparison of the absorption coefficients of silver bromide when dispersed in an emulsion and when in the form of a compact layer

Wavelength (Å)	Absorption coefficient (cm^{-1})		Ratio
	Emulsion layer	Compact layer	
3650	$2{\cdot}94 \times 10^4$	$0{\cdot}57 \times 10^4$	5·2
4047	$1{\cdot}52 \times 10^4$	$0{\cdot}162 \times 10^4$	9·4
4360	$1{\cdot}14 \times 10^4$	$0{\cdot}0484 \times 10^4$	23·5
4500	$0{\cdot}77 \times 10^4$	$0{\cdot}0251 \times 10^4$	31·5

Meidinger and Mucha (1939) have also drawn attention to the large numerical difference in absorption coefficient between granular and compact layers of silver bromide. In this connection measurements of Meidinger (1944) on the reflection and transmission of various pure bromide emulsions for the wavelength of 4360 Å are of interest. They are quoted in table 2 together with further information also given by Meidinger, and with derivations from his data to serve the purpose of this note.

It will be noted that the data for emulsions 1 to 5 yield results for the absorption coefficient ratio in dispersed and compact form of magnitude very similar to those

* Communication No. 1587H from the Kodak Research Laboratories.

obtained using the photographic method of measuring emulsion absorption described earlier. There is an exception in the case of the Lippmann emulsion, however, which is highly significant.

From inspection such emulsions are known to be practically non-scattering, which strongly suggests that the large increase of absorption coefficient as deduced from measurements on emulsions really represents a considerable increase of light path, and hence a higher fraction of net incident radiation absorbed in virtue of the light scattering properties of such layers.

An alternative explanation of the absorption increase is that it represents extra long wavelength absorption introduced by the very much larger fraction of surface bromide ions in the dispersed specimen. If this were true, by far the largest absorption coefficient ratio should be observed with the Lippmann

Table 2. Absorption of 4360 Å Radiation and other Properties of Pure Silver Bromide Emulsions

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	0·4	0·43	12	46	42	$0·77 \times 10^4$	15·9
2	0·3–1·8	0·32	13	36	51	$0·81 \times 10^4$	16·8
3	0·4–4·0	0·17	27	29	44	$0·49 \times 10^4$	10·2
4	1·6–5·5	0·13	31	24	45	$0·47 \times 10^4$	9·8
5	0·2	0·90	9	54	37	$0·78 \times 10^4$	16·1
6 (Lippmann)	0·03	6·0	77	12	11	$0·115 \times 10^4$	2·4

(1) Emulsion number; (2) grain size (μ); (3) mol % of ions on grain surface; (4) transmission (%); (5) reflection (%); (6) absorption (%); (7) absorption coefficient; (8) ratio of disperse to compact absorption coefficient.

emulsion. In fact the reverse is the case. Thus one is forced to conclude that scattering is responsible for the large differences in apparent absorption coefficient.

Strong confirmation of this view is provided by the pronounced wavelength dependence of the ratio as illustrated in table 1. The ratio decreases as the wavelength decreases and the true absorption coefficient increases; this is to be expected if scattering is responsible for the ratio since the higher the fraction of incident radiation absorbed by a grain, the smaller the fraction remaining to be scattered.

Hence it would seem reasonable to conclude from experimental evidence that the scattering of light in turbid photographic layers can result in an increase of density to visible wavelengths by a factor of 20 or so. This does not mean that the rate of absorption of energy by a layer is increased by a factor of 20; under such scattering conditions the latter factor is only about 4 (Pitts 1954).

References

Meidinger, W., 1944, *see* Stenger, E., and Staude, H. (Editors). *Fortschritte der Photographie* (Leipzig: Akademische Verlagsgesellschaft), Vol. 3, pp. 31–44.

Meidinger, W., and Mucha, H., 1939, *Phys. Z.*, **40**, 239.

Pitts, E., 1954, *Proc. Phys. Soc.* B, **67**, 105.

Slade, R. E., and Toy, F. C., 1920, *Proc. Roy. Soc.* A, **97**, 181.

Webb, J. H., 1948, *J. Opt. Soc. Amer.*, **38**, 27.

REVIEWS OF BOOKS

Vision Through the Atmosphere, by W. E. K. MIDDLETON. Pp. xiv+250. (Toronto: University Press; Oxford: University Press, 1952.) 68*s*.

This book supersedes an earlier one by the same author, which, under the title *Visibility in Meteorology*, has become a standard text on visual range, the atmospheric and visual factors which determine it and the measurements by which it can be predicted. Some material dealing with weather forecasting and the collection of meteorological data has been omitted, but accounts of a great many recent investigations on visibility—most of them made during World War II in connection with Service problems—have increased the book to about double its previous size. This new work includes Blackwell's statistical study of visual thresholds, the determinations of scattering and absorption coefficients at various altitudes by Waldram, Dessens' direct observation of the growth of haze particles, experimental verifications of Koschmieder's formula by Coleman and others and a group of investigations by British workers on the visual range of objects in searchlight beams. The author has pruned away some dead wood, he has amplified the sections on photometric theory and throughout has clarified and improved the argument. The net result is highly satisfactory and the success of *Visibility in Meteorology* should be repeated. One of the most interesting chapters gives a revised account of the apparent colours of distant objects and the effect of colour on their visual range—questions whose elucidation we owe very largely to the author. Although the conclusion is drawn that "for practical purposes no special theory of the visual range of coloured objects is necessary" an adequate enquiry into the question like that given here is certainly important.

Koschmieder's theory of the effect of the atmosphere on the contrast between object and background continues to be the basis of most work on vision through the atmosphere. Duntley's attempt at a wider theory (two-constant theory) using the system of ideas first introduced by Schuster (1905) in discussing the distribution of radiation in stellar atmospheres, is subjected to the radical criticism that although formally it appears to give the correct result (where comparison with Koschmieder is possible) it does so in fact only if the absorption and scattering coefficients are interpreted as referring to collimated light and not to diffuse light as intended by Duntley. The author's criticisms of Duntley's theory seem to the reviewer fully justified.

On one point the author's discussion of Koschmieder's formula is equivocal. He says on page 107 that the simple exponential law for the reduction of contrast C, with distance R, ($C_R = C_0 e^{-\sigma R}$, where σ is the coefficient of attenuation), holds only if there is no true absorption, i.e. if the attenuation arises solely from scattering (see also the comment on formula 4.9, p. 64). This however is not the case, as may readily be seen, for example, from the very simple derivation of Koschmieder's formula which I gave in my review of the second edition of *Visibility in Meteorology* (*Proc. Phys. Soc.*, 1942, **54**, 72). From the context it is probable that on page 107 the author has in mind the determination of C_0, which is more difficult if the atmosphere absorbs as well as scatters. But this is a different question and does not affect the validity of the formula quoted. The

latter certainly involves assumptions which are satisfied for example for a uniform flat earth of indefinite extent with a stratified atmosphere uniform in properties in any plane parallel to the ground. But in each stratum the absorption and scattering coefficients may have any constant values.

In the few pages on the special problems of the meteorologist, the author discusses the official definition of what is to be taken as visible when viewing meteorological marks. The official definition (Conf. of Directors, Intern. Meteor. Organ., Washington 1947, Resolution No. 147) plumps for recognition of the object for what it is, a tree, for example. The several objections to this definition and the superiority of ' mere detection ' as the criterion of visibility are convincingly expounded by Dr. Middleton. The fact that ' recognition ' could be chosen at all, illustrates the need in many quarters for a clearer grasp of the problems treated in this book. W. S. S.

Colloque sur la Sensibilité des Crystaux et des Émulsions Photographiques, Paris, 1951, *Mémoires et Discussions*. Pp. xii + 380. (Paris: Éditions de la Revue d'Optique théorique et instrumentale, 1953.)

In September 1951, the fourth of the recent series of photographic conferences was held in Paris, following meetings in preceding years at Liége, Zurich and Bristol. These conferences, with their pleasantly informal atmosphere, attracted large numbers of research workers from the photographic industry and the universities all over the world, mainly from amongst those who are interested in the fundamental problems of photographic sensitivity, and in the properties of nuclear research emulsions.

The present volume contains all the papers submitted to the conference, and much of the discussions which followed the presentation of the papers. The number of communications was large, sixty-eight, and the range of subjects covered extended from crystal physics via the properties of the latent image and the mechanism of development to the photographic aspects of work with nuclear research emulsions—a wide range indeed, giving a good indication of the state of non-secret research into these subjects at the time of the conference.

All the papers are printed in French, original or translation as the case may be, and whether or not this is a good thing is very much a matter of personal taste. Whatever one's opinion, however, the question of language is not such an important one here, for this is not a book for the private bookshelf: it is a collection of papers which should find its place in the reference libraries both because of the intrinsic value of the matter it contains, and as a record of the state of fundamental photographic research in 1951. A. J. HERZ.

Régularité et séries divergentes, by PIERRE VERNOTTE. Pp. 52 + xi. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air (No. 282), 1953.) 800 fr.

Viruses in Medicine, various authors. Scientific Edition, FORREST FULTON. Vol. 9, No. 3 of *British Medical Bulletin*. Pp. iii + 76. (London: British Medical Bulletin, 1953.) 15*s*.

CONTENTS OF SECTION A

PAGE

Dr. J. G. Rutherglen, Dr. P. J. Grant, Dr. F. C. Flack and Mr. W. M. Deuchars. γ-Radiation from the Reaction $^{27}Al(p, \gamma)^{28}Si$ 101

Mr. J. S. Turner. Investigation of Triplet Neutron–Proton Scattering in the Low Energy Region using the Eckart and Bargmann Potentials . . . 111

Dr. G. N. Fowler. Electron–Neutrino Angular Correlation Functions in the Theory of Beta Decay 117

Dr. J. M. Cassels and Mr. J. D. Lawson. Absorption Cross Sections for 134 Mev Protons 125

Prof. S. Devons, Dr. G. Goldring and Dr. G. R. Lindsey. Emission of Electron–Positron Pairs from Light Nuclei—I: Monopole Transition in ^{16}O . . 134

Mr. R. H. Tredgold. A Modified Form of Heisenberg's Theory of Ferromagnetism 148

Dr. G. Harries. A Cloud Chamber Study of Internal Pairs from $^{12}C^*$ . . 153

Mr. B. A. Chartres and Prof. H. Messel. Three Dimensional Theory of Electron–Photon Showers 158

Mr. A. M. Lane and Dr. L. A. Radicati. Studies in Intermediate Coupling—II: Radiative Transitions in Light Nuclei 167

Lady Anne Pery. Isotope Shifts in the Atomic Spectrum of Calcium . . 181

Research Notes:

Dr. R. F. Barrow, Dr. G. Drummond and Dr. S. Walker. A Note on the Ultra-Violet Rand Spectra of CCl and SiCl 186

Mr. D. T. Stewart, Mr. P. W. F. Gribbon and Prof. K. G. Emeléus. Excitation Conditions for the Infra-Red Auroral Bands of Ionized Nitrogen 188

Dr. P. J. Jutsum. The Continuous Absorption of Light in Calcium Vapour . 190

Letters to the Editor:

Dr. E. Finlay-Freundlich. Red-Shifts in the Spectra of Celestial Bodies . 192

Dr. Max Born. On the Interpretation of Freundlich's Red-Shift Formula . 193

Dr. P. G. Klemens. The Thermal Conductivity of Monovalent Metals . 194

Mr. G. V. Marr. A Note on the Absorption of Light by Indium Vapour . 196

Reviews of Books 198

Contents of Section B 200

Negative Ion Emission from Oxide-Coated Cathodes: I

By F. A. VICK* and C. A. WALLEY†

* Physics Department, University College of North Staffordshire

† Electrical Engineering Department, King's College, Newcastle-upon-Tyne

MS. received 17*th November* 1953

Abstract. A simple glass-envelope mass spectrometer of the 90° deflection type is described. Its resolving power $\delta M/M$ is 0·00925, which is adequate for ions up to about $M = 50$. The dominant negative ions emitted from oxide cathodes are found in many cases to be chlorine, and it is shown that these often originate from the glass when baked to about 450°c, confirming the conclusions reached by Hamaker, Bruining and Aten. The chlorine ion emission is correlated with the poisoning of the cathode. Preliminary observations on other ions, $M = 16$, 26, 32, 42 and 44, are recorded and discussed.

§1. Introduction

The work described in this paper was commenced in 1946 as part of a programme on the fundamental physics of oxide cathodes and allied electron emitters. At that time it had been shown that the loss of sensitivity in use of a small region in the centre of the screen of magnetically focused cathode-ray tubes is due to bombardment by negative ions travelling with the electrons from the electron gun. Most of the published work describes experiments with sealed-off or demountable cathode-ray tubes to investigate the nature of these ions and methods of obviating 'ion burn' in the luminescent screens. Cathode-ray tubes were used as rudimentary mass spectroscopes, the various ions being deflected by an external magnet and detected by their effects on the luminescent screen (von Ardenne 1935, Levy and West 1936, Bachman and Carnahan 1938, Broadway and Pearce 1939). Apparently a great variety of ions was present under different conditions, including those of mass-to-charge ratios, M/e, of 1, 8, 12, 16, 18, 19, 26, 30, 32, 35, 37, 40, 42, 56, 74, 80, 101, 120, 127, 169, 261. Some of these disappeared as the oxide cathodes aged, indicating that they probably originated in the binder. Only rough estimates of relative ion currents could be made.

As is well known, further interest in the mechanism of electron emission from oxide cathodes has been stimulated by their extended use during the war, especially for pulse work, and by the realization that the growing understanding of the nature of semiconductors might be applied with advantage to such cathodes. It seemed likely that ion movement within the oxides during and after activation might be of fundamental importance. It therefore appeared worth while to investigate, from the point of view of the cathode, and not of the luminescent screen, the emission of negative ions from oxide cathodes at various stages of activation and life, and especially to study quantitatively the emission of oxygen ions. Preliminary experiments were therefore undertaken with a simple form of mass spectrograph

specially designed for the purpose. It was soon found that if proper precautions were taken to obtain a good vacuum and freedom from impurities, only a few varieties of ions were emitted from the cathode, and that normally the predominant ones were chlorine 35 and 37. An investigation was made into the origin of these ions and it was eventually found that they came directly or indirectly from the glass. This was confirmed independently and by quite different methods by Hamaker, Bruining and Aten (1947). and has been verified since by others. A preliminary survey was made of the scope of the mass spectrograph as an aid to the understanding of the structure and properties of oxide cathodes, and the results are given below. During the investigation further papers became available (Schaefer and Walcher 1943, Evans 1947, Sloane and Watt 1948) and their results are compared with those obtained in the present investigation. Further work based on these investigations will be described in additional papers.

§2. Apparatus

To avoid contamination as far as possible it was decided to use an all-glass mass spectrometer evacuated by a mercury diffusion pump through liquid-air traps. The normal sequence of cleaning, baking and pumping was followed. In the earlier investigations, especially when studying ion emission from the cathode during and immediately following activation, the spectrometers were continuously pumped, but provision was made for sealing off and using getters.

All spectrometers so far made for this investigation have been of the 90° deflection type, and have been designed for use with an existing magnet giving a field of up to 7000 oersteds over a circular pole face 2 inches in diameter. The relationship between ion mass M, accelerating potential V volts, magnetic field H oersteds, the ion path in the magnetic field l cm and the angle θ radians through which the ions are deflected is

$$\theta = 0{\cdot}69 \times 10^{-2} Hl(1/MV)^{1/2}. \qquad \ldots\ldots(1)$$

Neglecting the fringing field and assuming the field between the 2 in. circular poles to be uniform, the maximum length of ion path in the field is $2{\cdot}54\pi/\sqrt{2}$ cm. For $\theta = 90°$ and a field of 7000 oersteds the maximum value of MV which can be used is 3×10^4. If a collecting slit 1 mm wide is placed 15 cm from where the beam leaves the pole pieces, the resolving power $\delta M/M$ is 0·00925. For a mass number of 40, $\delta M = 0{\cdot}37$, which is adequate, since it was intended to investigate the emission of ions in the neighbourhood of $M = 32$.

The first spectrometers incorporated electron gun assemblies designed for standard VCR 97 cathode-ray tubes.* The electron guns consist of cathode, modulator or grid, first anode, focusing electrode and final anode. The two anodes are kept at the same potential so as to avoid changes in beam focus with variations in grid potential. Various constructions were tried. The one used for most of the experimental work is shown in figure 1. The button cathode supports are quite independent of the remainder of the electron gun. By cracking the glass of the side tube, 1, the cathode assembly can be removed, the cathode itself replaced, the whole reassembled and the side-tube re-sealed without disturbing the rest of the electron gun structure, 5. In this way the focusing and alignment of the ion beam can remain the same for a series of experiments. The whole

* We are much indebted to Cinema Television Ltd. of London S.E.16 for the provision of gun assemblies.

spectromenter is of hard glass (Hysil). The glass wall between the final node of the electron gun and the entrance slit of the collector is coated with 'Aquadag' (grade 660 B, which has no organic binder) which is carefully baked on to the glass. This coating is at the same potential as the final anode of the accelerating gun and the entrance slit of the collector. The central curved section of the envelope, radius 10 cm, is flattened for insertion between the magnet pole pieces. The collector system consists of an entrance slit in a nickel plate at the end of a nickel cylinder connected to the Aquadag coating and to the anode, and a collector electrode within the Faraday cylinder.

The ion currents to the collector pass to earth through a high resistance of the order of 10^{10} ohms, the potential difference across which is fed to the grid of an electrometer tetrode valve (Ferranti BM 2). The whole head amplifier so formed is carefully shielded and insulated and placed very close to the collector electrode,

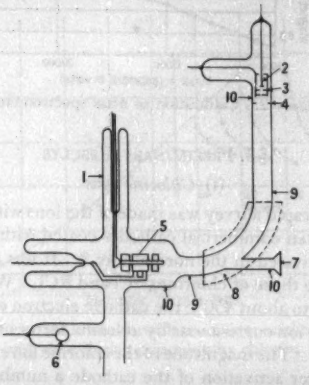


Figure 1. The glass mass spectrometer. 1 Cathode support tube, 2 Collector, 3 Cylinder and entrance slit, 4 Baffle, 5 Electron gun, 6 Penning gauge, 7 Window, 8 Region of magnetic field, 9 Limits of flattening, 10 Limits of graphite coating.

with a direct lead from collector to electrometer grid. The electrometer anode current is suitably amplified and passed to a galvanometer, with the usual precautions for stabilization. Ion currents down to 10^{-14} amp can be measured. Secondary electron emission at the collector can be suppressed by maintaining it at about 20 volts positive with respect to the entrance slit.

The gas pressure in the spectrometer is monitored, especially during activation of the cathode, by a Philips-type gauge* capable of measuring pressures down to 10^{-7} mm Hg or less.

The ion spectrum is swept across the entrance slit by slow and continuous variation of electromagnet current, using a substantial wire-wound variable resistance with a motor-driven contact remotely controlled. (A liquid resistance with moving plate was tried, but found not to be sufficiently steady.) At the same time the galvanometer currents are recorded, the rate of variation of magnetic field being sufficiently slow for the time constants of the electrometer and galvanometer circuits not to cause appreciable error.

* Description to be submitted for publication.

The spectrometer was calibrated by two methods, (*a*) by the introduction of a known gas through a controlled leak and ionizing it with a tungsten filament in the cathode position, (*b*) by the use of existing ^{35}Cl and ^{37}Cl ions (see below). In this case the accelerating potential was varied in steps, and the magnet current determined for the positions of the ^{35}Cl and ^{37}Cl peaks in the mass spectrum curves for each value of potential. A plot of MV (mass × voltage) against magnet current then gave a smooth curve from which the mass of any other ion within the range may be determined, assuming that the ions are singly charged (see equation (1) above). A typical calibration curve is shown in figure 2.

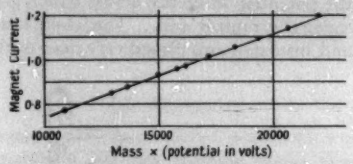


Figure 2. Calibration of mass spectrometer.

§ 3. Preliminary Results

(i) *Chlorine Ions*

A preliminary and rapid survey was made of the ions with mass numbers up to about 45 using about ten commercial cathodes coated with suspensions of mixed $BaCO_3$ and $SrCO_3$ activated in the normal way.* It was at once found that the predominant ions were those of chlorine, ^{35}Cl and ^{37}Cl. When the glass spectrometer tube was baked to about 450°C the cathode electron emission fell and at the same time the chlorine ion current rose by at least a factor of 10 compared with the current before baking. The magnitude of the chlorine ion emission then gradually fell. Immediately after activation of the cathode a number of other ions were found in addition, in comparatively small quantities, but these soon disappeared (in a few hours) and were attributed to the remains of the organic binder in the original suspension. In some cases ions of M/e 26, 32, 42 and 43 persisted with the chlorine ions. There was no reason to expect that the chlorine ions came from anything but impurities, and since their presence was an obvious complication in the study of ion emission associated with fundamental oxide-cathode behaviour, the source of the chlorine was sought.

In a further series of experiments great care was taken to eliminate chloride impurities from cathode materials, electrodes etc., but in spite of all the precautions, including the use of spectroscopically pure materials in the cathode suspension, chlorine ions were always found. At first they increased in abundance when a new cathode was brought into use, and then decreased gradually during tens or hundreds of hours. In view of the observation noted above, the effects of baking the glass on the poisoning of the cathode and on the magnitude of the chlorine ion currents were investigated in more detail. It was confirmed that baking the Hysil to about 400–450°C led to poisoning of the cathode after it had been running

* We are indebted to the Research Laboratories of the General Electric Co., Wembley, to Ferranti Ltd. Manchester, and to Cinema Television Ltd. of London for gifts of suspensions.

for 2 to 45 minutes. Some dependence of ion current on modulator (grid) potential was also observed. The circuit was then modified to permit the measurement of electron currents between cathode and modulator for various modulator potentials. When this potential was raised to 8 volts positive with respect to the cathode, the electron current began to decay with time. The experimental procedure was to reduce the modulator–cathode voltage to zero and then to raise the heater voltage to 5·5 or so for one minute. The heater voltage was then reduced to the standard 4·5 and the cathode–modulator volts increased to 4. If the electron current was then within 5% of the standard value, the modulator potential was suddenly raised to the required value, and the variation of modulator current with time determined. Typical results are shown in figure 3, indicating that poisoning commences at about 8 volts, and that as the modulator potential is increased above this value the rate of poisoning is increased.

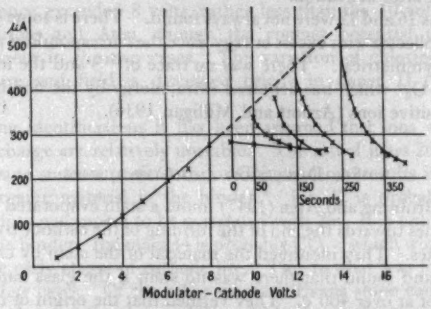


Figure 3. Variation of electron emission with modulator–cathode potential difference, and with time during poisoning by chlorine ions.

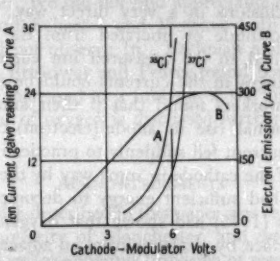


Figure 4. Variation of chlorine ion current with modulator potential, and poisoning of cathode.

These results were correlated with chlorine ion emission by plotting ^{35}Cl or ^{37}Cl ion currents received at the collector of the spectrometer against modulator potential. Figure 4 shows the kind of curve obtained. It is clear that poisoning occurs when the chlorine ion currents begin to increase rapidly. Some complexities in the ion spectrum were noticed, especially when the modulator voltage was in the region of 6 or 7.

It was also shown that baking the glass envelope increased (though not immediately) the chlorine ion currents, indicating again that some at least of the chlorine originated in the glass and that the poisoning observed was related to the appearance of this chlorine. From the positions and shapes of the ion peaks it was inferred that some chlorine ions were from the cathode and some from the modulator.

(ii) *Other Ions*

Of the other ions detected, that of mass 26 was shown to vary in amount with the presence or absence of organic binder in the cathode. The currents of ions of mass 42 and 44 were very small, except with some of the commercial cathodes examined in early experiments. The current of ion mass 42 did not change with the amount of binder present. Ion 44 was noticed in one or two cases only, and then only soon after activation.

Ions of mass 16 and 32 were not always found. There is some evidence that 32 arrives at the collector with excess energy since the corresponding peak in the mass spectrum is asymmetrical. There was no trace of ^{34}S and the ion of mass 32 is assumed to be O_2^- which has acquired extra energy by the bombardment of the cathode by positive ions (Arnott and Milligan 1936).

§4. Discussion and Conclusion

Hamaker, Bruining and Aten (1947) found a stain evaporated on to the anode surfaces of diodes towards the end of the forming of the cathode by decomposition of the carbonates. They identified the material of the stain by chemical analysis as a chloride, and found that there was no stain if the glass had been baked at 200°c instead of at over 400°c. They verified that the origin of the chloride was the glass by showing that no stain was formed in a tube constructed of specially made chlorine-free glass, even when baked at over 400°c. Our results confirm these results and conclusions in a very direct way, since the chlorine ions evaporated from the cathode or liberated from the modulator by electron bombardment contributed to the measured ion currents, the ions could be identified, and the changes in the currents could be related to other relevant factors. The Dutch workers found that if their anode voltage was steadily raised there was a normal rise in anode (electron) current until the voltage reached 10, when the current fell suddenly to practically zero. They attributed this to the poisoning of the cathode in some way by the chlorine liberated when the incident electrons had sufficient energy to decompose the chloride on the anode surface. Jacobs (1946) had shown that oxygen may be liberated from oxides on an anode surface by electrons from an adjacent cathode, and that the oxygen may cause poisoning of the cathode. Plumlee and Smith (1950) have confirmed the evaporation of both BaO and $BaCl_2$ from an oxide cathode at temperatures above 600°c, using a mass spectrometer in which evaporated atoms and molecules were subsequently ionized by an electron beam. Bachman and Carnahan, and Sloane and Watt, using sealed-off tubes, found that after considerable use only ions of mass 35 and 37 persisted with appreciable intensity, but Schaefer and Walcher reported that after a few hours 35 and 37 disappeared though other ions persisted. These latter workers, however, used a demountable apparatus with greased joints which could not be baked, so their

observations are not inconsistent with the conclusions reached above. Plumlee and Smith found that the largest negative ion peaks were those of the chlorine isotopes, and that the ion currents were both pressure and temperature dependent. They conclude that the chlorine ions are evaporated thermally from the cathode, and suggest that the cathode collects neutral chlorine compounds or atoms and ejects negative chlorine atoms. They found, confirming our observations, that the chlorine ion currents were sensitive to the torching of the tube at various points.

Our results indicate clearly the connection between chlorine ion current and poisoning of the cathode. The experiments confirm that the chlorine originates mainly in the glass, that ions are emitted thermally from the cathode and that chlorides evaporated on to the modulator are decomposed by electron bombardment so as to emit more chlorine ions. In our case appreciable poisoning of the cathode emission was observed when the cathode-modulator potential difference exceeded 8 volts, rather less than the 10 volts reported by Hamaker, Bruining and Aten, though the contact potential difference may have been different in the two cases. The variation of chlorine ion emission with temperature and field is discussed briefly in paper II (Grattidge and Shepherd 1954).

In suggesting identifications it has been assumed that ions with more than one electronic charge are relatively unstable. The ion of mass 26 could be CN^- or $C_2H_2^-$. Sloane states a preference for CN^-, and our results indicate that its origin is the organic material in the binder.* Ion 42 is probably CNO^-, and its abundance does not appear to be correlated with 26. It is unlikely, therefore, to come from the binder. Ion mass 44 is probably CO_2^-, which would be expected only during the early stages of cathode life. This is what was found, and even then only in a few cases. Of the other workers, Evans alone commented on its presence. Plumlee and Smith found 26 and 14 generated at the cathode surface, probably by a positive ion bombardment mechanism, since there was little temperature dependence. We did not find appreciable quantities of 14.

Sloane and Watt and Schaefer and Walcher both reported the presence of 32, but the latter did not observe 16. Bachman and Carnahan obtained 16 but not 32. Evans, and Broadway and Pearce found both. Plumlee and Smith report O_{32}^- ion current increased as a funcion of the emission drawn from the cathode. The behaviour of oxygen is discussed further in paper II (Grattidge and Shepherd 1954).

Acknowledgments

The experimental work described in this paper was carried out in the Physical Laboratories of the University of Manchester, and we are much indebted to Professor P. M. S. Blackett for generous provision of research facilities. F. A. V. gratefully acknowledges the help of a Grant for Special Research from the Department of Scientific and Industrial Research, and C. A. W., first of a maintenance grant from the Department of Scientific and Industrial Research, and then of a Research Fellowship from Associated Electrical Industries Ltd. Acknowledgment is also made to Mr. F. Rowerth for technical assistance and to our colleagues, especially Dr. W. Grattidge and Dr. A. A. Shepherd, for helpful discussion.

* *Note added in proof.* Bachman, Hall and Silberg (1953) prefer $C_2H_3^-$.

References

VON ARDENNE, M., 1935, *Arch. Elektrotech.*, **29**, 731.
ARNOT, F. L., and MILLIGAN, J. C., 1936, *Proc. Roy. Soc.* A, **156**, 538.
BACHMAN, C. H., and CARNAHAN, D. W., 1938, *Proc. Inst. Radio Engrs, N.Y.*, **26**, 529.
BACHMAN, C. H., HALL, G. L., and SILBERG, P. A., 1953, *J. Appl. Phys.*, **24**, 427.
BROADWAY, L. F., and PEARCE, A. F., 1939, *Proc. Phys. Soc.*, **51**, 335.
EVANS, W. S., 1947, *R.A.E. Technical Note R.A.D.* 389.
GRATTIDGE, W., and SHEPHERD, A. A., 1954, *Proc. Phys. Soc.* B, **67**, 177.
HAMAKER, H. C., BRUINING, H., and ATEN, A. H. W., Jr., 1947, *Philips Res. Rep.*, **2**, 171.
JACOBS, H., 1946, *J. Appl. Phys.*, **17**, 596.
LEVY, L., and WEST, D. W., 1936, *J. Instn Elect. Engrs*, **79**, 11.
PLUMLEE, R. H., and SMITH, L. P., 1950, *J. Appl. Phys.*, **21**, 811.
SCHAEFER, H., and WALCHER, W., 1943, *Z. Phys.*, **121**, 679.
SLOANE, R. H., and WATT, C. S., 1948, *Proc. Phys. Soc.*, **61**, 217.

Negative Ion Emission from Oxide-Coated Cathodes: II

By W. GRATTIDGE* and A. A. SHEPHERD†

* Department of Physics, Massachusetts Institute of Technology, now at the General Electric Co. Ltd., Schenectady, U.S.A.
† Department of Physics, University College of North Staffordshire.

Communicated by F. A. Vick; MS. received 17th November 1953

Abstract. The present paper describes a detailed study of negative ion emission processes from oxide cathodes during all stages of cathode life, using two types of mass spectrometer having oxide cathode sources. The main ions found were those of mass numbers 12, 16, 26, 27, 32, 35, 37, 42 and 43. A retarding potential method was employed to determine energy distributions of all the ions, and special attention was paid to ions emitted during activation, it being found that any oxygen evolved during activation was in an uncharged state. Variations in ion emission due to cathode temperature and applied field were also studied.

§ 1. Introduction

The experiments described in the present paper were performed as a continuation of the work described in the paper "Negative Ion Emission from Oxide-Coated Cathodes—I" (Vick and Walley 1954, to be referred to as I). Although previous mass spectrometer studies of ion emission from oxide cathodes existed (notably those of Schaefer and Walcher (1943), and Sloane and Watt (1948)), no attempt had been made to correlate the ion emission with the state of activity of the cathode and the stages of cathode life. The object of this investigation was to distinguish between such stages of electron emission activity and to seek correlations with the ion emission and also to study the temperature and accelerating field dependence of the ion emission.

The previous work has shown that the principal negative ions found in tubes operating with oxide cathodes have M/e values of 16, 26, 32, 35, 37, 42, 43 and the possible sources are (*a*) ions emitted thermionically from the cathode, (*b*) ions formed at the grid or modulator by electron dissociation, (*c*) ions ejected from the surface of the cathode by back bombardment of positive ions formed by ionization of the residual gas in the tube. The masses of the ions are obtained by use of a mass spectrometer and the origins of the ions are determined by a retarding potential method.

§ 2. Apparatus

The experimental work was performed using a small mass spectrometer apparatus developed from the type of instrument described in I for investigating ions of low masses (up to mass 45). Two different types of glass experimental tubes as shown in figure 1, both using 90° deflection, were operated. Figure 1(*a*) shows the initial tube which used a commercial type electron gun to focus the beam of electrons and negative ions emitted from the cathode. The electrode

arrangement consisted of cathode, modulator or grid, first accelerating anode, focusing electrode and final anode. The first accelerating anode was operated at the same potential as the final anode and enabled the grid voltage to be varied between very wide limits without causing any defocusing of the emergent beam.

Figure 1(*b*) shows the simpler tube designed as a direction-focusing mass spectrometer. The electron gun of this tube consisted of a cathode, grid and anode.

Both tubes were made of Hysil glass in the forms shown. Three sections of the tubes can be distinguished: (*a*) the source, consisting of the cathode and electron gun, (*b*) the deflection chamber, (*c*) the collector for the measurement of ions appropriately deflected.

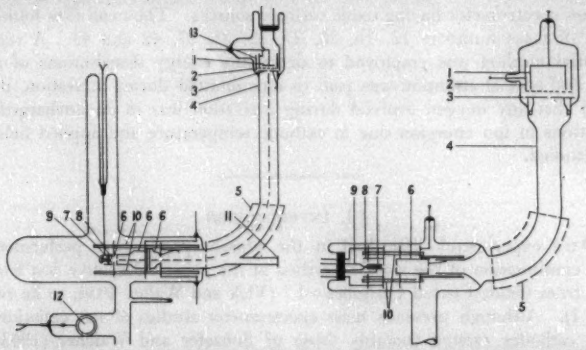


Figure 1. The two forms of mass spectrometer tube used in the investigation. (*a*) employed a commercial electrode system for focusing, (*b*) used the focusing property of the direction-focusing mass spectrometer.

1 collector, 2 retarder, 3 exit slit, 4 graphited tube, 5 region of magnetic field, 6 anodes, 7 cathode, 8 modulator, 9 heater, 10 ceramic rods, 11 window for alignment, 12 penning gauge, 13 earthed screen.

(*a*) The sources were arranged so that both cathodes and grids were demountable and could be easily changed without affecting the alignment of the gun. In the direction-focusing spectrometer the cathode and grid were changed as one unit, the alignment being maintained by the ceramic rods attached to the anode structure (see figure 1(*b*)). The cathodes were of area 0·04 cm^2 with a cathode–grid spacing of approximately 0·2 mm. The cathodes were coated with a commercial barium–strontium carbonate mixture. The temperatures of the cathodes were determined by an optical pyrometer sighted either through the viewing window or from the side of the cathode. In both tubes the grid was operated positive with respect to the cathode for higher beam currents.

(*b*) The deflection chamber (of radius 10 cm and 7 cm respectively in the two tubes) fits between the poles of the electromagnet. The inner surface of the glasswork of this section was coated with Aquadag.

(*c*) The collector heads for both tubes were very similar and consisted of a slit in a nickel plate maintained at anode potential and connected to the graphite coating, a nickel gauze to act as a retarding electrode, and a collector. In the

case of the second tube the condition for direction focusing is that the ion source (the cathode), the centre of curvature of the magnetic field, and the collector slit should be collinear (Inghram 1948).

The collector consisted of a Faraday cylinder connected by a short lead to the input resistance of a d.c. amplifier used to measure the collector currents.

The tubes were continuously evacuated by a mercury diffusion pump with liquid-air trap and rigorous baking and outgassing schedules were applied. The pressure in the system (10^{-7} mm Hg) was monitored by a Philips type gauge.

The magnet was movable so that the tubes could be baked to 450°c and metal parts outgassed by radio-frequency heating.

The mass spectrum was scanned by continuous variation of the magnetic field, and for constant accelerating voltages up to 1 kv and a field up to 2000 gauss masses up to 45 could be scanned. Figure 2 shows a typical plot of the complete spectrum showing the resolution of the instrument. Both forms of spectrometer were capable of similar resolution. The direction-focusing tube had the advantage that there was less metal work in the tube and outgassing was much easier.

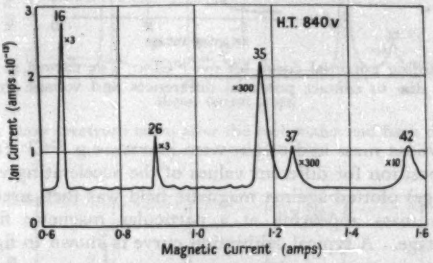


Figure 2. A typical mass spectrum showing the resolution obtainable. (Multiplying factors to be used to obtain true peak height.)

The ion currents were measured using a d.c. amplifier built around a directly heated single electrometer tube (Ferranti BM2) and a 10^{11} ohm grid resistor. The minimum ion current detectable was 1×10^{-14} amp.

§ 3. Retarding Potential Measurements

The collecting system described in the previous section included a retarding electrode consisting of a nickel gauze placed between the defining slit and the Faraday cage. By applying a variable voltage between this electrode and the cathode, a retarding field region could be established in front of the collector and the energy distribution of the collected ions investigated by a study of the collector current as a function of this retarding voltage. Figure 3 shows a typical retarding plot for $^{35}Cl^-$.

§ 4. Experimental Procedure

All metal parts used in the assembly of the tubes were hydrogen fired before use and outgassed by radio-frequency heating on the system where possible. The glass work was first chemically cleaned, then the deflection chamber was coated with Aquadag colloidal suspension, dried and baked to 400°c in air.

After assembly each tube was sealed in position on the vacuum system, evacuated and baked out to 450°c for several hours, and on cooling a pressure of the order of 10^{-7} mm Hg was obtained.

The mass spectra were scanned using a slowly decreasing magnetic field for constant ion accelerating voltage. Calibration was carried out using a

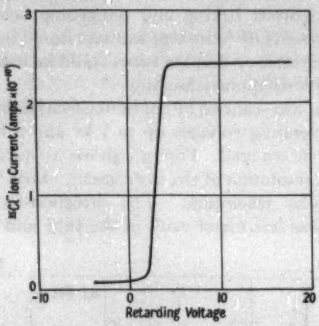


Figure 3. A retarding potential curve for ion $^{35}Cl^-$. The cut-off at 2 v positive of the cathode is due to contact potential differences and voltage drop in the cathode coating.

particular ion whose mass and origin were known, e.g. $^{35}Cl^-$ from the cathode, and noting its position for different values of the accelerating voltage. A graph of (mass × voltage) plotted against magnetic field was then used to identify any other unknown mass appearing at a particular magnetic field for a given accelerating voltage. A typical calibration curve is shown in figure 4.

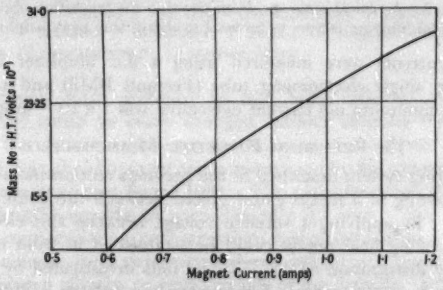


Figure 4. A typical calibration curve for the mass spectrometer.

§5. Results

(i) *Ions Emitted on Activation*

Each cathode was outgassed and 'formed' in a rigorous manner so that the vacuum in the system was good before activation proceeded and spectra were recorded. This involved operating the cathode at a temperature not greater

than 1100°K for a period of one or two hours until all the gases evolved during the decomposition of the alkaline earth carbonates had been pumped away and the vacuum was better than 10^{-6} mm Hg. A short activation flash (1250°K) was then given and since little gas was evolved, spectra could be recorded from the initial stages of cathode life.

Figures 5 and 6 show two typical spectra taken during the processing of a cathode. Figure 5 was taken after the cathode had been heated at a temperature of about 1100°K for two hours and for a grid voltage of 10 V the emission current was about 37 mA cm^{-2}. The ions recorded were those of mass numbers 12, 16, 26, 27, 35, 37, 42 and 43.

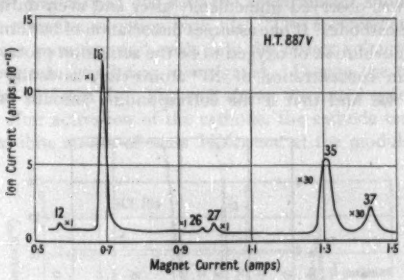


Figure 5. A mass spectrum taken after the carbonates had been decomposed at a temperature of 1100°K.

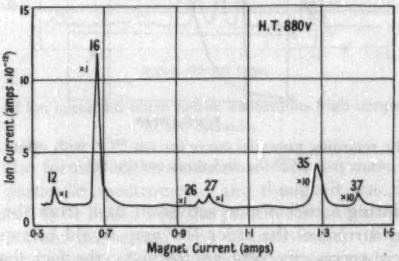


Figure 6. A mass spectrum taken after flashing the cathode to 1250°K.

The ion spectrum with the cathode at 1100°K after flashing is shown in figure 6, when the emission current density was increased to 78 mA cm^{-2}. There is no alteration in intensity of the ions of mass numbers 12, 16, 26 and 27. The chlorine ion intensity, however, is diminished by a factor of approximately four. This is assumed to be due to preferential evaporation of the cathode surface layers of barium chloride at the high flashing temperature. The formation of barium and strontium chlorides by the action on the cathode of hydrochloric acid gas released on bake-out of glass apparatus is discussed in I.

Retarding potential measurements performed on the ions indicated that ions of mass numbers 12, 16, 26, 27, 35, 37 and 42 had their origin in the

region of the cathode. In the chlorine ion peaks the distributions (figure 3) show a cut-off at a point 2 to 3 volts positive of the cathode, which is assumed to be due to contact potential differences, and the voltage drop in the cathode coating. Only the ion of mass 16 exhibits any excess energy. An excess energy 'tail' extending to about 10 volts in excess of the accelerating voltage may be seen on its distribution (figure 7).

The investigation of the oxygen ion emission shows that the activation process, although it probably leads to evolution of oxygen during the production of excess barium in the coating, does not give rise to the emission of ionized oxygen from the cathode. In all cases no appreciable increase of the oxygen 16 peak height was observed immediately after and even during the activation flash given to the cathode. If one assumes dissociation of barium oxide producing free barium with evolution of oxygen to be the activation process, then assuming an excess barium concentration of 10^{18} atoms/cm^3, as found by Jenkins and Newton (1949), we find that if the corresponding amount of ionized oxygen

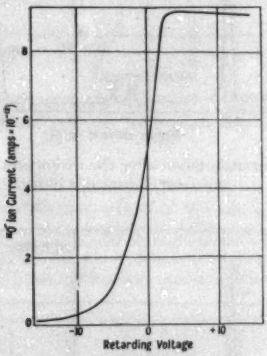


Figure 7. The retarding potential curve for ion $^{16}O^-$ with origin at the cathode showing an excess energy tail.

were liberated during a two-minute activation flash from the cathode coatings used then an ion current of the order 10^{-7} amp would be expected. Although this calculation is necessarily very approximate, the fact that no oxygen ion currents greater than 10^{-11} amp were found on activation gives a direct confirmation of the conclusions of Becker (1929), Detels (1927) and Isensee (1937) that most of the oxygen emitted when current is drawn is uncharged.

It seems probable that the oxygen ions arise from the disintegration of a surface oxygen film on the cathode by positive ion bombardment. This is supported by the fact that in the present investigation, where the total accelerating voltage was never more than 900 v and the vacuum of the order of 10^{-7} mm Hg, only a small amount of ionized oxygen was evolved, whereas in the work of Sloane and Watt, where the gas pressure in the apparatus was rather higher, and an accelerating voltage up to 2800 v used (thus giving greater probability of high-energy positive-ion bombardment of the cathode) the oxygen ion intensity in relation to that of the other ions was greater. Also, the oxygen

ion 16 found by Sloane and Watt exhibited considerable excess energy (up to 100 v) while in the present work the excess energy 'tail' of this ion was only 10 v in length.

A similar explanation would also account for the excess energy exhibited by ions of mass number 1, 24, 25, 26 and 32 as found by Sloane and Watt, since it has been shown (Arnot and Milligan 1936, Sloane and Press 1938), that energy in excess of the energy of bombarding ions can be given to particles liberated by such bombardment.

(ii) *Ions Emitted during Life*

An experiment was carried out with the spectrometer shown in figure 1(*a*) to determine ion emission variation with life. The spectrometer was operated without continuous pumping and with only the heater voltage applied, except when the ion spectrum was being investigated. During the test the pressure remained at about 10^{-6} mm Hg. Figure 8 shows the initial spectrum of the test taken 160 hours after activation of the cathode, the cathode temperature being 1250°K. This exhibits an ion of mass 34 formed at the modulator and a small

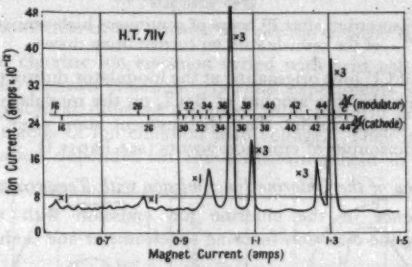


Figure 8. Initial ion spectrum taken before continuous high temperature running of the cathode.

peak of mass 16 also formed at the modulator. The ion mass of 34 was always found with this particular spectrometer, and though it was not identified, its occurrence appeared to be linked with the presence of a ceramic cathode support which was absent from the second type of instrument. Figure 9 shows the spectrum of the ions emitted after the cathode was operated continuously for a period of 70 hours at a temperature of 1250°K. A more complex spectrum resulted, with the 16, 42 and 43 mass ions having increased, and an additional peak of mass number 26·7 formed at the modulator. The latter was identified as mass 27, Al^-, which was also found by Sloane and Watt, and which is thought to arise from the use of an alumina heater coating. The oxygen ions of mass number 16 with origin at the modulator appear to be formed by electron dissociation of barium oxide evaporated from the cathode at this high temperature. The chlorine ion intensity had increased by a factor of more than two during this test period. This behaviour may be compared with that described in (i) when a reduction in Cl^- intensity was found after a short flash to 1250°K, due to the disturbance of the low temperature equilibrium. The gradual growth and eventual decay of the Cl^- ions on running for a prolonged period

at 1200°K is presumably due to the new conditions of equilibrium between formation and evaporation of barium chloride which are set up at this higher temperature. Both the ion intensity and the electron emission became insignificant after 200 hours of this high temperature treatment.

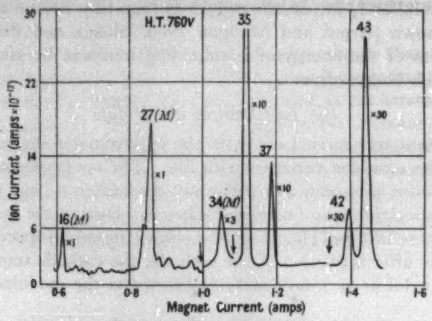


Figure 9. Ion spectrum after 70 hours of continuous high temperature running of the cathode with no current being drawn.

The absence of Cl^- ions originating at the modulator during these experiments appears to indicate that the amount of $BaCl_2$ on the modulator after 160 hours running is very much smaller than that present soon after activation, when serious chlorine poisoning of emission occurs (see Paper I).

(iii) *Variation of the Chlorine Ion Emission with Temperature and Field*

The dependence of the chlorine ion emission with temperature was investigated using the direction-focusing spectrometer and is shown in figure 10.

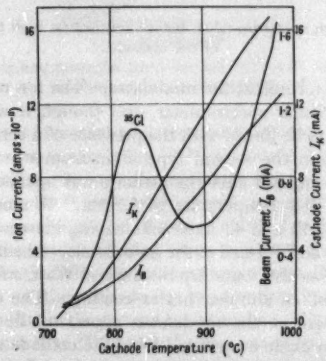


Figure 10. Dependence of chlorine ion emission on temperature. I_K and I_B indicate corresponding variations of cathode and beam currents.

This shows the existence of a maximum and minimum of chlorine ion intensity in the region of 800°–900°c. This phenomenon was also observed by Sloane and Watt, who did not study it in detail but who found it repeatable. Figure 10

also includes the corresponding cathode and beam current variations with temperature.

Experiments were also made to investigate the dependence of chlorine ion emission on accelerating field. Earlier measurements shown in figure 11

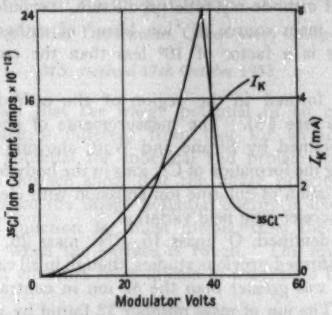


Figure 11. Variation of chlorine ion emission and electron emission I_K with grid voltage.

indicated that the chlorine ion emission varied with grid voltage in a manner different from the electron emission. Several experiments were carried out for the cathode placed at different distances from the grid and observing the corresponding variation of ion emission with voltage. They indicated that the chlorine ion emission depended on the accelerating field in rather a critical

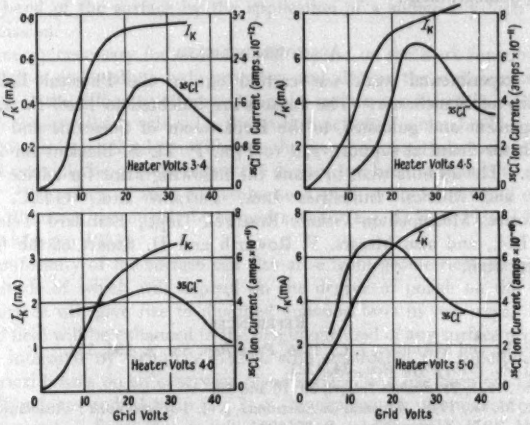


Figure 12. The variation of chlorine ion emission with grid voltage at several cathode temperatures.

manner, showing a fairly well defined maximum. A group of graphs showing the variation of electron and chlorine ion emission with grid volts at various cathode temperatures is shown in figure 12. Other ions behaved in a similar way in contrast to the dependence of the electron emission.

§ 6. Conclusions

(1) It is confirmed that in the emission of negative ions from oxide cathodes the predominant ions are those of the chlorine isotopes of masses 35 and 37. Since the conditions in this investigation approximated closely to those appertaining in industrial cathode-ray tube production, it would seem probable that these ions are the main source of 'ion burn' of cathode-ray tubes. Their maximum intensity is a factor of 10^6 less than the corresponding electron emission.

These ions are formed in the region of the cathode from barium and strontium chlorides (see § 5). The measurements of energy distribution are similar to those obtained by Sloane and Watt, showing an energy spread of 1–2 volts, suggesting the formation of Cl^- ions in the body of the cathode coating. There was no correlation of chlorine ion emission with electron emission either by temperature or accelerating field variation.

(2) Other ions identified, O^- mass 16, CN^- mass 26, Al^- mass 27, CNO^- mass 42, and 43, confirmed previous studies, though in all cases here the intensity of the mass 43 ion was greater than the 42 ion in contrast to the findings of Sloane and Watt. The ion of mass number 32 found by Sloane and Watt and others only occurred in one case after prolonged running of the cathode, and was too small to permit detailed measurements of its energy distribution.

(3) Most of the oxygen evolved during activation of the cathode appears to be emitted as uncharged molecules. However, during life singly charged oxygen ions appear to be produced in a tube having a small grid–cathode separation by dissociation of barium oxide (evaporated from the cathode on the modulator) by electron bombardment.

Acknowledgments

The experimental work was carried out in the Physical Laboratories, University of Manchester. The authors are indebted to Prof. F. A. Vick for encouragement and guidance, to the Department of Scientific and Industrial Research for financial support, and to Prof. P. M. S. Blackett for laboratory facilities. The authors wish to thank the following firms for advice and help: Electric and Musical Industries Ltd., Ferranti Ltd., G.E.C. Research Laboratories, Metropolitan-Vickers Research Dept., Standard Telephone & Cables Ltd., and also Messrs. F. Rowerth and H. Stuart of the University Technical Staff.

References

Arnot, F. L., and Milligan, J. C., 1936, *Proc. Roy. Soc.* A, **156,** 538.
Becker, J. A., 1929, *Phys. Rev.*, **34,** 1323.
Detels, F., 1927, *Jb. drahtl. Telegr.*, **30,** 10 and 52.
Inghram, M. G., 1948, *Advances in Electronics*, Vol. 1 (New York: Academic Press).
Isensee, H., 1937, *Z. phys. Chem.*, B, **35,** 309.
Jenkins, R. O., and Newton, R. H. C., 1949, *J. Sci. Instrum.*, **26,** 172.
Schaefer, H., and Walcher, W., 1943, *Z. Phys.*, **121,** 679.
Sloane, R. H., and Press, R., 1938, *Nature, Lond.*, **159,** 302.
Sloane, R. H., and Watt, C. S., 1948, *Proc. Phys. Soc.*, **61,** 217.
Vick, F. A., and Walley, C. A., 1954, *Proc. Phys. Soc.* B, **67,** 169.

The Work Function of Irregular Metal Surfaces

By T. J. LEWIS
Electrical Engineering Department, Queen Mary College, London

MS. received 17*th October* 1953

Abstract. It is shown that the image potential of an electron outside an irregular metal surface is different from that for an ideal plane surface. Calculations of this potential for spherical and prolate spheroidal bosses on an otherwise smooth surface show that the image potential might easily be reduced by 50%. Since the image potential energy is likely to form a large portion of the work function for most metals, then the latter also will be significantly reduced when the surface is rough. The importance of this in the theory of electron emission from metals in practice is briefly discussed.

§ 1. Introduction

Before an electron is emitted from the surface of a metal it has to acquire sufficient energy to overcome the binding forces of the metal. Such energy may be acquired thermally by raising the temperature of the metal, as in thermionic emission, or by the incidence of an energetic particle or photon, or it may be emitted as a result of a modification of the potential barrier in the neighbourhood of the surface by the application of a sufficiently high field as in cold emission.

This energy necessary for emission will depend on the work function of the metal, which is essentially a bulk property, and also on the surface conditions.

Because of the difficulty of controlling the surface conditions it is extremely difficult to obtain reliable values for the work function of a metal. The effective work function is immediately modified by the presence of surface contaminations which will certainly be present unless the surface has been subjected to rigorous cleaning. The effects of such contaminations on thermionic and cold emission have been discussed thoroughly by several authors (e.g. de Boer 1935, Llewellyn Jones and Morgan 1953).

Non-uniformity of the surface can also arise from geometrical irregularities, the magnitude of which will depend on the degree of polish on the surface. Such a surface will give rise to modified emission laws in two ways. Firstly, an applied field will be enhanced in the neighbourhood of any surface projection. This was indicated by Schottky (1923), who considered the superposition of hemispherical bosses on an otherwise plane surface, and has been used recently by Llewellyn Jones and Morgan (1953) in the explanation of emission processes from cold metals. Secondly, as will be shown subsequently, there will be a lowering of the energy necessary for emission, i.e. a lowering of the effective work function, by an amount which will depend on the geometry of the projection.

It is the purpose of the present work to consider these variations in the effective work function due to the micro-geometry of the surface as distinct from those variations which result from adsorbed layers of foreign materials. In order

that the problem should be amenable to numerical calculation it will be necessary to consider certain simple types of projection only. From the results obtained, however, it should be possible to indicate the behaviour for a much wider range of projections and to discuss the likely variation of the work function for surfaces occurring in practice. The nature of the work function for a perfectly smooth surface must be considered first since the range of validity of the classical treatment to be used must be known.

§ 2. The Work Function of a Smooth Surface

The work function ϕ of a metal is defined in terms of the change in energy of a neutral metal lattice when an electron from the Fermi level is completely removed. In the process of removal, forces due to the positive ions of the lattice and to the remaining electrons give rise to a potential which will vary from inside to outside the metal. The difference of potential energy of an electron inside and outside the metal is then W_a and

$$W_a = W_l + W_p + W_e + W_s$$

where W_l and W_p are contributions due to ionic and polarization effects of the lattice and W_e and W_s are the combined exchange and correlation energies and the surface layer energy contributions respectively.

Within the metal, since the electron considered is in a region of almost maximum allowable electron density, the contributions from W_l and W_p will not enter the problem (Slater 1934). Outside the metal W_l and W_p combine and give rise to the classical image potential due to the change induced on the surface of the metal by the electron. This potential energy has the value $-e^2/4x$ at a distance x from a perfectly plane surface.

When the electron is within the metal it will exist in a hole created by the action of mutual repulsion forces between the electrons. These forces are of two kinds. Firstly, the operation of the exclusion principle for electrons of the same spin produces the so-called Fermi hole. The resulting energy is the exchange energy, the magnitude of which will depend on the momentum of the electron under consideration. Secondly, repulsion of an electrostatic nature gives rise to a correlation energy which, however, is small. Bardeen (1936) has accounted for this contribution by an addition to the exchange energy giving a resultant energy W_e.

Although within the metal W_e is constant it will become less negative in the neighbourhood of the surface, due to an increase in the size of the Fermi hole. In this region also a surface double layer, giving energy W_s, will occur due to the fact that the electron distribution will not be symmetrical about the surface ions.

The combined potential energies W_e and W_s will merge smoothly into the image potential at a small distance from the surface of the metal (Slater 1934, Bardeen 1936).

Calculations of the variation of these energies through the surface have been made for sodium by Bardeen (1936), and figure 1, taken from his paper, shows the combined potential energy for an electron having the maximum Fermi energy as a function of distance. It is found that the classical image force fails at a distance x_0 from the surface, where x_0 is about 2 Å. Sachs and Dexter (1950) have also indicated that failure of the image force may be expected at about this distance. The calculation of Bardeen is for monovalent sodium, but it would be

reasonable to assume that the general conclusions, but not the details, will be the same for other metals.

If E_m is the maximum kinetic energy of the electron corresponding to the Fermi level, then the work function ϕ is given by

$$e\phi = W_0 + W_s - E_m$$

where e is the charge on the electron.

For sodium ϕ has the calculated value 2·35 v, of which about 80% is contributed by the image force. Since no information is available concerning the form or magnitude of the combined energy for other metals, it is impossible to estimate the proportion of their work functions to be ascribed to image forces. It would seem likely however that image calculations would cease to be valid at about the same distance (2 Å) from the surface.

The failure of the image law at small distances from the surface was realized at an early date by Schottky (1923) and Langmuir (1916). Schottky assumed that at distances less than x_0 the forces were constant, while Langmuir assumed

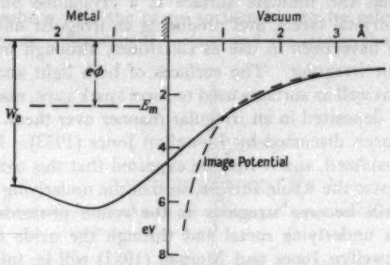


Figure 1. Combined potential energy of an electron having the maximum Fermi energy E_m.

a parabolic law merging smoothly into the image law at the distance x_0. The estimated work function was then $e/2x_0$, and x_0, calculated from observed values of the work function, was a few angström units for most metals. While the estimates of x_0 obtained by Schottky and Langmuir are of the right order, their assumptions now appear very arbitrary.

The particular form of the image potential $e/4x_0$ arises because an ideal plane boundary surface to the metal is assumed. If, as is most likely in practice, the surface is irregular, then the form of the image potential will change according to the surface geometry. If these irregularities are of sufficient size (perhaps at least an order or so greater than atomic dimensions) then the exchange and correlation forces which are essentially short range effects could be considered unchanged and the alteration in work function for such an irregular surface estimated from the modification of the image potential alone.

The modification to the classical image potential when the surface is irregular will now be considered. As already stated, only simple geometrical irregularities are amenable to calculation, but these can be made to cover quite a range of possible irregularities and will indicate the general behaviour. It will be assumed

that the classical laws will hold up a to distance of 2 Å from the surface, which, from the work of Bardeen, appears reasonable.

Before proceeding with this problem, it is pertinent to consider the degree of surface irregularity to be expected in practice.

§ 3. The Surface Micro-Geometry

The micro-geometry of surfaces used in practice as electron emitters is likely to be far from plane. Surfaces which have been mechanically polished will have surface markings of a degree dependent on the abrasive material used in polishing. While it is true that such polishing will tend to produce amorphous layers, it is not unlikely that the final surface will exhibit the underlying irregularities in many cases. Electrolytic polishing is also likely to leave the surface in an irregular state (Evans and Hopkins 1952). Processes such as grinding or turning will, most certainly, produce considerable roughness with projections of up to 10^{-3} cm in height (Bikerman 1944).

Even in cases where a high degree of smoothness has been obtained, it must be remembered that the ultimate surface is a crystalline one consisting of variously oriented crystal facets, and producing an irregular micro-geometry.

Surfaces which have been in use as electrodes, although initially polished, will tend to become irregular. The surfaces of both light and heavy current electrical contacts, as well as surfaces used to form spark gaps, may become pitted or may have metal deposited in an irregular manner over them. The nature of such surfaces has been discussed by Llewellyn Jones (1953). Most frequently these surfaces are oxidized, and it may be expected that this oxidation does not proceed uniformly over the whole surface, so that the underlying metal, although initially smooth, will become irregular as the action proceeds. Emission of electrons from this underlying metal and through the oxide film as recently investigated by Llewellyn Jones and Morgan (1953) will be influenced by this micro-geometry and the corresponding variation in the effective work function.

Although only cold surfaces have been mentioned it is also true that hot surfaces such as those for thermionic emission may be rough. A filament which has been polished so that its surface has an amorphous layer will tend to lose this as the temperature approaches the melting point. Further, at such temperatures, the application of high fields will cause the surface atoms to pile up according to the local field direction (Benjamin and Jenkins 1940), and this may cause severe irregularity of the surface.

To summarize, it may be said that projections having heights in the range 10^{-7} to 10^{-5} cm might easily occur on the surfaces of emitters used in practice. This range will be investigated in the following.

§ 4. The Image Potential of Irregular Surfaces

The contribution to the effective work function from the classical image potential will be calculated for certain idealized surface irregularities, assuming that the classical law breaks down at a distance $x_0 (\sim 2\,\text{Å})$ from the surface. This will be compared with the corresponding contribution $e^2/4x_0$ which is the value when the surface is ideally smooth.

The quantity to be calculated is the potential energy of the electron at a distance x_0 from the surface due to the electrostatic forces of the charges induced on the surface by the electron considered. The induced charges alone will

produce a field outside the metal and the magnitude of the potential of a point charge of magnitude e in such a field at a distance x_0 will be twice the potential energy required.

The calculations will be for single projections on an otherwise smooth surface, since only in such cases is a treatment possible. Two such systems will be considered: (*a*) The projection is in the form of a spherical boss which is immersed in the surface by varying amounts. The limiting cases are those in which the sphere is just making contact with the surface and where it is completely immersed to give a perfectly plane surface. An intermediate case occurs where the sphere is sufficiently immersed to form a hemispherical boss. (*b*) Starting with the hemispherical boss a further system may be developed consisting of halves of prolate spheroids on the otherwise flat surface. This system includes pointed projections from the surface of varying degree of sharpness.

4.1. *Spherical Projections*

The spherical system is shown in figure 2, in which the sphere has a radius b. The potential energy of an electron at P where QP is x_0 and O is the point of contact of the sphere with the plane surface has to be calculated. The plane and sphere form an equipotential surface and the image charges in this surface

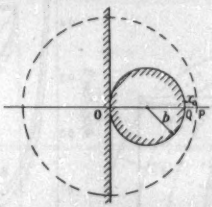


Figure 2. Spherical boss on a plane surface.

induced by the electron at P form an infinite set. The contributions to the potential at P due to this set have to be summed and it is better to invert the system before doing so. The circle of inversion is chosen with centre O and radius OP. As shown in Appendix I, the inverted system is that of two parallel planes with the electron at a point which is not equidistant from the planes. It is then possible to sum the potentials in a way indicated by Maxwell (1881) for a similar problem.

On inverting back again to the original system, the potential at P due to the image charges then becomes

$$V_P = e\left[\frac{1}{2x_0} - \frac{b}{(2b+x_0)^2}\left\{2\gamma + \Psi\left(\frac{x_0}{2b+x_0}\right) + \Psi\left(1-\frac{x_0}{2b+x_0}\right)\right\}\right]$$

where γ is Euler's constant 0·5771 and $\Psi(x)$ is the digamma function $(d/dx)\log\Gamma(1+x)$, which has been tabulated by Lodge and Wishart (1931). Thus the potential energy of the electron at P becomes $-\theta_1 e^2/4x_0$ where

$$\theta_1 = 1 - \frac{2bx_0}{(2b+x_0)^2}\left\{2\gamma + \Psi\left(\frac{x_0}{2b+x_0}\right) + \Psi\left(1-\frac{x_0}{2b+x_0}\right)\right\}$$

is the correction factor to the normal image law potential for a perfectly plane surface and $\theta_1 \to 1$ as $b \to 0$ or ∞.

Values of θ_1 for various values of b are shown in figure 3 with x_0 equal to 2 Å. By plotting θ_1/x_0 it is possible to compare the variation of the potential with distance from the surface with the corresponding potential for a plane surface, and this has been done in figure 4 for representative values of b.

It will be seen that, for b less than about 40 Å, θ_1 decreases rapidly, and for projections with radii in this range, a significant reduction in the effective work function might be expected. When b is less than about 1 Å, θ_1 increases again, but values of the radius which are less than 5 Å are probably outside the range of the present treatment.

4.2. *Hemispherical Boss*

This problem of the hemispherical boss is similar to that for the spherical boss and is considered in Appendix I. The potential energy now becomes $-\theta_2 e^2/4x_0$ where

$$\theta_2 = x_0/(b+x_0) + 4b^3x_0/\{(b+x_0)^4 - b^4\} \qquad \ldots\ldots(1)$$

and again $\theta_2 \to 1$ and $b \to 0$ or ∞.

Values of θ_2 are shown in figure 3 and the variation of θ_2/x_0 with distance is shown in figure 4.

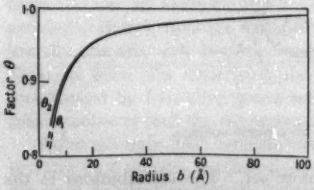


Figure 3. Variation of θ_1 and θ_2 with the boss radius b of a spherical projection.

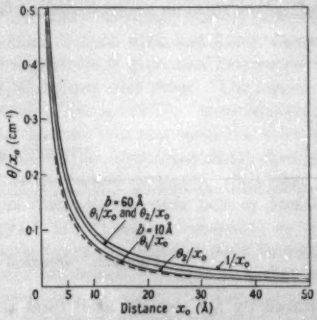


Figure 4. Comparison of θ_1/x_0 and θ_2/x_0 with x_0^{-1}, the value for a plane surface.

It will be seen that θ_1 and θ_2 are practically equal when b is greater than 10 Å. Below this value θ_2 first decreases more slowly than θ_1 and finally $\theta_2 \to \theta_1 \to 1$ when $b \to 0$.

Since, when the sphere is completely immersed in the surface, the normal image potential for a flat surface will be obtained, the curves of figure 5 have been sketched. These indicate the likely variation of θ, the correction to the normal image law, with the degree of immersion of the spherical projection in the surface for various values of b. There is little difference between the spherical and hemispherical boss when the radius is greater than about 10 Å. If the lower limit to b is taken at 5 Å, then a 15% reduction in the image component results for both cases.

4.3. *Prolate Spheroidal Projections*

The hemispherical boss is the limiting case of the prolate spheroidal type of boss. Such a boss may be represented as one of a family of spheroidal surfaces

as shown in figure 6. In this system the confocal spheroids (η constant) represent equipotential surfaces while the orthogonal hyperboloids (ξ constant) are lines of force. A spheroidal boss on an otherwise plane surface then becomes the natural boundary $\eta=\eta_1$, $\xi>0$; $\xi=0$, $\eta>\eta_1$. The height of the boss b and the base radius a are the semi-major and semi-minor axes of the spheroid, and as a is decreased with b fixed, the boss becomes more and more needle-like in shape. Since $\eta_1=b/(b^2-a^2)^{1/2}$, $\eta_1\to 1$ as $a\to 0$. Thus the height and sharpness of the projection may be varied by alteration of b and η_1.

The image potential for an electron at P($\xi=1$, $\eta=\eta_0$) on the axis of such a system, where QP (figure 6) is the distance x_0 from the point at which the electrostatic treatment is no longer applicable, can be found in terms of prolate spheroidal harmonics involving the Legendre polynomials P_n and Q_n.

It is shown in Appendix II that the potential at P due to the induced charge in the metal alone is

$$V_P = e\sum_{n=0}^{\infty}\frac{(2n+1)P_n(\eta_1)Q_n^2(\eta_0)}{(b^2-a^2)^{1/2}Q_n(\eta_1)}$$

and thus the potential energy of an electron at P is given by

$$E_P = -e^2/2(b^2-a^2)^{1/2}\sum_{n=0}^{\infty}\frac{(2n+1)P_n(\eta_1)Q_n^2(\eta_0)}{Q_n(\eta_1)}. \qquad \ldots\ldots(2)$$

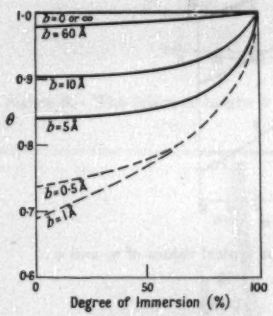


Figure 5. The dependence of θ on the degree of immersion of a spherical boss in a plane surface.

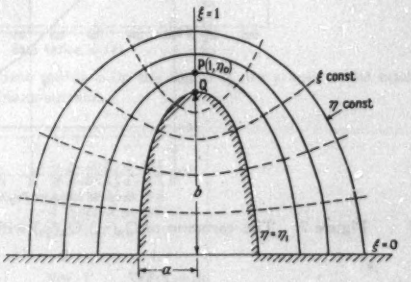


Figure 6. Prolate spheroidal boss on a plane surface.

This sum cannot be evaluated easily but it is not difficult to show how the expression behaves for various values of η_1. It will be shown that the sum is always less than a certain expression which can be evaluated so that an upper limit to the potential energy will be obtained. It will be seen that for many shapes of projection the energy will be well below this limit.

Now $Q_n(\eta)$ may be expressed in terms of the hypergeometric function as

$$Q_n(\eta)=\frac{\sqrt{\pi}\Gamma(n+1)}{2^{n+1}\Gamma(n+\frac{3}{2})\eta^{n+1}}F(\tfrac{1}{2}n+\tfrac{1}{2},\ \tfrac{1}{2}n+1;\ n+\tfrac{3}{2};\ \eta^{-2})$$

and it can be shown that

$$F(\tfrac{1}{2},1;\tfrac{3}{2};\eta_1^{-2})/F(\tfrac{1}{2}n+\tfrac{1}{2},\tfrac{1}{2}n+1;n+\tfrac{3}{2};\eta_1^{-2}) < F(\tfrac{1}{2},1;\tfrac{3}{2};\eta_0^{-2})/F(\tfrac{1}{2}n+\tfrac{1}{2},\tfrac{1}{2}n+1;n+\tfrac{3}{2};\eta_0^{-2})$$

and therefore, since $\eta_0 > \eta_1 \geqslant 1$,

$$Q_0(\eta_1)/Q_n(\eta_1) \leqslant Q_0(\eta_0)/Q_n(\eta_0)$$

or

$$Q_n(\eta_0)/Q_n(\eta_1) \leqslant Q_0(\eta_0)/Q_0(\eta_1) = \ln\frac{\eta_0+1}{\eta_0-1}\Big/\ln\frac{\eta_1+1}{\eta_1-1}.$$

Thus $(2n+1)Q_n^2(\eta_0)P_n(\eta_1)/Q_n(\eta_1) \leqslant (2n+1)P_n(\eta_1)Q_n(\eta_0)\ln\frac{\eta_0+1}{\eta_0-1}\Big/\ln\frac{\eta_1+1}{\eta_1-1}$.

Now $P_n(\eta_1)Q_n^2(\eta_0)/Q_n(\eta_1)$ and $P_n(\eta_1)Q_n(\eta_0)$ are positive for $n \geqslant 0$ and $\sum_0^\infty (2n+1)P_n(\eta_1)Q_n(\eta_0)$ converges and has the value $(\eta_0-\eta_1)^{-1}$ (Hobson 1931), so that the expression (2) for the potential energy becomes

$$|E_P| \leqslant e^2[2(b^2-a^2)^{1/2}(\eta_0-\eta_1)]^{-1}\ln\frac{\eta_0+1}{\eta_0-1}\Big/\ln\frac{\eta_1+1}{\eta_1-1}.$$

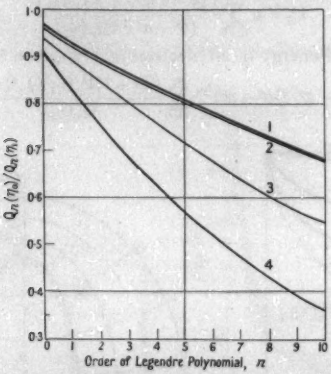


Figure 7. The variation of $Q_n(\eta_0)/Q_n(\eta_1)$ with n for typical values of η_0 and η_1.

	1	2	3	4
η_0	2·587	1·869	1·387	1·193
η_1	2·502	1·809	1·342	1·154

Writing $\eta_0 = (b+x_0)(b^2-a^2)^{-1/2}$; $\eta_1 = b(b^2-a^2)^{-1/2}$ gives

$$|E_P| \leqslant e^2 \ln\frac{\eta_0+1}{\eta_0-1}\Big/ 2x_0 \ln\frac{\eta_1+1}{\eta_1-1}$$

$$= \theta_3 e^2/4x_0 \quad \text{where} \quad \theta_3 = 2\ln\frac{\eta_0+1}{\eta_0-1}\Big/\ln\frac{\eta_1+1}{\eta_1-1}. \quad \ldots\ldots(3)$$

For a fixed height of the projection b both η_0 and η_1 decrease from infinity to unity as the base radius a decreases from the value b to zero. As can be seen from figure 7, θ_3, which equals $2Q_0(\eta_0)/Q_0(\eta_1)$, also decreases. This figure also shows that the ratio $Q_n(\eta_0)/Q_n(\eta_1)$ decreases rapidly as n decreases, especially when η_0 and η_1 are small, and therefore the inequality expressed by (3) will also increase as n increases.

Thus $|E_P|$ must be less than $(\theta_3 e^2/4x_0)_{\eta_1=\infty}$ for $\eta_1 = \infty$ and must then decrease more rapidly than $\theta_3 e^2/4x_0$ as $\eta_1 \to 1$.

Now $\eta_1 = \infty$ gives the limiting case of a hemispherical boss for which E_P has been given exactly by eqn (1) as equal to $-\theta_2 e^2/4x_0$, and thus $|E_P|$ may be written as $|E_P| \leqslant \beta\theta_3 e^2/4x_0$ where $\beta = \theta_2/(\theta_3)_{\eta_1=\infty}$.

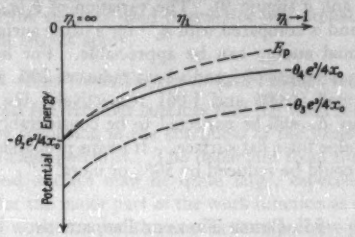


Figure 8. Sketch showing the probable dependence of E_P on the parameter η_1.

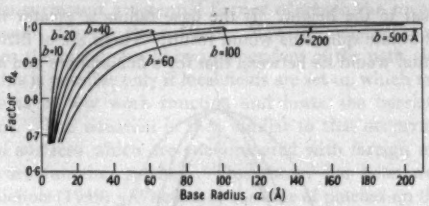


Figure 9. The effect of height b and base radius a on the factor θ_4 for a spheroidal boss on a plane surface.

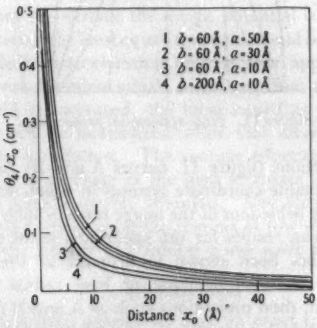


Figure 10. Comparison of θ_4/x_0 with $1/x_0$.

The value of θ_3 for $\eta_1 = \infty$ may be evaluated from the expressions for η_0 and η_1 whence $\beta = \theta_2(b+x_0)/2b$, and therefore

$$E_P \geqslant -\theta_2\theta_3 \frac{b+x_0}{2b}\frac{e^2}{4x_0} = -\frac{\theta_4 e^2}{4x_0} \qquad \ldots\ldots (4)$$

where $\theta_4 = \theta_2\theta_3(b+x_0)/2b$. The value of E_P must agree with eqn (1) for $\eta_1 = \infty$ and must then increase more rapidly than the relation (4) for η_1 decreasing

(figure 8). It may be expected, therefore, that the actual value of E_P will vary in some such manner as that shown in figure 8.

Making use of the values of θ_2 obtained earlier, θ_4 has been computed for various values of b and a (figure 9). The variation of θ_4/x_0 with distance x_0 is shown in figure 10 and is compared with x_0^{-1} for a plane surface. The reduction in the image potential energy can be appreciable. For instance, θ_4 has the value 0·89 for a boss of height 60 Å and a base radius of 20 Å, and since the values of η_0 and η_1 are small (1·096 and 1·061 respectively) the magnitude of the corresponding energy E_P will be expected to be considerably less than 89% of the corresponding value for a flat surface. It seems probable that in many cases the image potential could be reduced by 50% or more.

§ 5. Other Types of Projection

The cases treated so far have been projections in which the base radius was less than or equal to the height. If the base radius is greater than the height then a family of oblate spheroids would result. These are of little interest since the image potential would lie between that for a hemisphere and a plane surface.

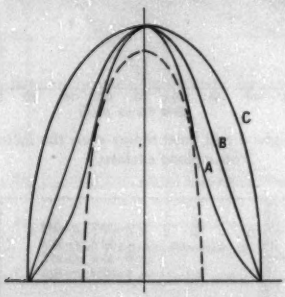


Figure 11. Non-spheroidal projections.

More severe projections (figure 11, curves A and B) have not been treated. It is difficult to find suitable coordinate systems in which such points are natural boundaries. The likely behaviour of the image force in such cases can be assessed however from the known results for the spheroidal point (figure 11, curve C). The image potential has been shown to decrease as the base radius of the projection decreases. Since a decrease of base radius effectively means a sharpening of the point, then projections such as A and B (figure 11) which are sharper than the spheroidal projection C should have lower image potentials. In fact the type of projection A will have an image potential which is close to that of the inscribed prolate spheroid of much smaller base radius.

The present treatment has been for single projections on an otherwise smooth surface, but a normal surface, of course, will be covered with projections of various magnitudes, depending on the history of the surface. These irregularities may take the form of ridges, serrated along the top, interspersed with projections of the type already considered. Provided individual projections stand out above the general level of irregularity of the crystal surface, the results

given here will indicate likely reductions of the image potential. If, however, there are no individual projections above the general level of the crystal surface then it is to be expected that the reduction in the image potential will be small. The majority of the surfaces used in practice will probably have projections to which the present results apply.

§ 6. The Work Function and Patch Fields

It has been shown for irregular surfaces that the work function of individual projections will vary according to the sharpness of the projection on account of the change in the image potential. The departure from the value of the work function for an ideal surface may be quite large, especially when the image potential accounts for the major part of the work function as in sodium.

If the calculated work function thus varies from point to point on the surface, an electron emerging from within the metal and moving out to be infinitely distant would appear to surmount a potential barrier of which the magnitude depends on the path chosen through the surface. Now the amount of work to extract an electron from the metal should be independent of the path chosen, and for a rough surface this is possible only if local fields are set up which raise the effective barrier in regions of low work function and lower the barrier in high work function regions. The situation is then similar to that occurring in the patch theory of metal surfaces which are contaminated with foreign materials to give low or high work function patches. This theory has been well reviewed by Herring and Nichols (1949). Whereas in the case of patches on the metal surface the patch fields extend out to large distances from the surface, depending on the size of the patches, it would appear that the local fields set up by geometrical irregularities will have a much shorter range. It will be seen from figures 4 and 10 that at about 50 Å from the surface the image potential is very nearly that for a plane surface and therefore the local equalizing field would be small at that distance.

If an external, electron-accelerating field is now applied, the surface will exhibit an average work function which will depend on the field strength. In particular, as the field is increased, the local retarding field above low work function projections will be neutralized while that above high work function projections will not be so affected. The average effective work function of the surface will thus tend to decrease as the field increases until at high fields the local retarding fields are completely annulled. Any electron emission at such fields will come almost entirely from the regions of low work function, i.e. from the projections, and the surface will exhibit a low work function.

§ 7. Conclusions

Although it has not been possible to give an exact treatment for the image potential of an irregular surface, the approximate treatment has indicated that for surfaces where a moderate degree of roughness is likely there will be a significant reduction in this part of the work function. The remainder of the work function due to exchange forces, etc., should not be greatly affected by surface roughness since these forces are essentially short range ones. In view of the approximate nature of our argument, the overall reduction in the work function ϕ due to irregularities cannot be estimated with any precision. It would not seem surprising however if individual projections had work functions which were at least 25% less than that for a plane surface.

It should be noted also that for projections where the reduction in work function is greatest, the enhancement of any applied field is also greatest. This double effect will be important for high field thermionic emission (Schottky 1923) and particularly so for cold field emission in which the emission current i is given approximately by an expression of the form $i = AF^2 \exp(-B\phi/F)$ where F is the field in the neighbourhood of the emitting area and A and B are constants.

Projections for which the work function ϕ is low and the corresponding local applied field F is large will make an overwhelming contribution to the total emission current of a given surface. Any estimate of the current density on the basis of the emission occurring uniformly from the whole surface area may then be seriously incorrect.

ACKNOWLEDGMENTS

The author wishes to thank Dr. H. Tropper for helpful discussions on the paper and Professor W. J. John for facilities provided in the department. The work has been done during the author's tenure of the W. T. Henley Fellowship which is gratefully acknowledged.

REFERENCES

BARDEEN, J., 1936, *Phys. Rev.*, **49**, 653.
BENJAMIN, M., and JENKINS, R. O., 1940, *Proc. Roy. Soc.* A, **176**, 264.
BIKERMAN, J. J., 1944, *Rev. Mod. Phys.*, **16**, 53.
DE BOER, J. H., 1935, *Electron Emission and Adsorption Phenomena* (Cambridge : University Press).
EVANS, D. J., and HOPKINS, M. R., 1952, *J. Electrodep. Tech. Soc.*, **28**, paper 8.
HOBSON, E. W., 1931, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge : University Press).
HERRING, C., and NICHOLS, M. H., 1949, *Rev. Mod. Phys.*, **21**, 185.
LANGMUIR, I., 1916, *Trans. Amer. Electrochem. Soc.*, **29**, 157.
LLEWELLYN JONES, F., 1953, *J. Instn Elect. Engrs*, Pt. I, **100**, 169.
LLEWELLYN JONES, F., and MORGAN, C. G., 1953, *Proc. Roy. Soc.* A, **218**, 88.
LODGE, A., and WISHART, J., 1931, *Mathematical Tables*, Vol. I (London : British Association).
MAXWELL, J. C., 1881, *Electricity and Magnetism* (Oxford : Clarendon Press).
SACHS, R. G., and DEXTER, D. L., 1950, *J. Appl. Phys.*, **21**, 1304.
SCHOTTKY, W., 1923, *Z. Phys.*, **14**, 63.
SLATER, J. C., 1934, *Rev. Mod. Phys.*, **6**, 209.

APPENDIX I

(i) *Image Potential for a Sphere on a Plane Surface*

The problem is to calculate the potential energy of a point charge $-e$ at x_0 (figure 2) in the presence of its image charges in the surface which is at zero potential.

Since the image charges form an infinite set, the problem is better treated by inverting the system about the centre O and with a radius of inversion $OP(=r)$. The inverted system (figure 12) consists of a charge $-e'$ at P′ between two infinite parallel planes at zero potential. The image charges in this system are located as follows:

charges $-e'$ at distances $\pm 2n(r+c)$

charges $+e'$ at distances $2n(r+c)-2r$ and $-[2n(r+c)-2c]$

where $c = r^2/2b - r$ and distances are measured from P′; $n = 1, 2, \ldots . \infty$.

These images will produce a potential V_P' at P' where

$$V_P' = -\sum_1^\infty \frac{e'}{n(r+c)} + \tfrac{1}{2}\sum_1^\infty \frac{e'}{n(r+c)-r} + \tfrac{1}{2}\sum_1^\infty \frac{e'}{n(r+c)-c}$$

$$= -\tfrac{1}{2}e'\sum_1^\infty \frac{r}{n(r+c)[n(r+c)+r]} - \tfrac{1}{2}e'\sum_1^\infty \frac{c}{n(r+c)[n(r+c)+c]} + \frac{e'}{2}\left(\frac{1}{c}+\frac{1}{r}\right)$$

$$= -\frac{1}{2}\frac{e'}{r+c}\sum_1^\infty \frac{r/(r+c)}{n\{n+r/(r+c)\}} - \frac{1}{2}\frac{e'}{r+c}\sum_1^\infty \frac{c/(r+c)}{n\{n+c/(r+c)\}} + \frac{e'}{2}\frac{r+c}{rc}$$

$$= e'\left[\frac{r+c}{2rc} - \tfrac{1}{2}[\Phi\{r/(r+c)\}+\Phi\{c/(r+c)\}]/(r+c)\right]$$

where $$\Phi(x) = \sum_1^\infty \frac{x}{n(n+x)} = \gamma + \frac{d}{dx}\log\Gamma(1+x) = \gamma + \Psi(x)$$

in which γ is Euler's constant 0·5771 and $\Psi(x)$ is the digamma function which has been tabulated by Lodge and Wishart (1931).

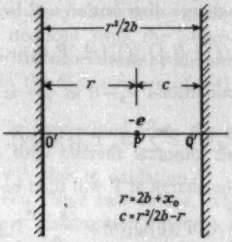


Figure 12. The inverted system for a spherical boss on a plane surface.

Inverting back into the original system gives $V_P' = V_P$ and $e' = e$, so that on substituting for Φ

$$V_P = e\left[\frac{r+c}{2rc} - \tfrac{1}{2}(2\gamma + \Psi\{r/(r+c)\} + \Psi\{c/(r+c)\})/(r+c)\right]$$

or $$V_P = e\left[1/2x_0 - \frac{b}{(2b+x_0)^2}[2\gamma + \Psi\{x_0/(2b+x_0)\} + \Psi\{1-x_0/(2b+x_0)\}]\right].$$

(ii) *Image Potential for a Hemisphere on a Plane Surface*

For the hemisphere only three image charges are produced, and these are situated on the axis joining the centre of the hemisphere O which lies in the plane surface with the position of the electron P. The charges are of magnitude e, $\{-b/(b+x_0)\}e$, $\{b/(b+x_0)\}e$ situated at distances $-(b+x_0)$, $-b^2/(b+x_0)$, $b^2/(b+x_0)$ from O respectively, where b is the radius of the hemisphere. The image potential will then be

$$V_P = \frac{e}{2(b+x_0)} = \frac{eb}{(b+x_0)^2+b^2} + \frac{eb}{(b+x_0)^2-b^2}$$

which may be written as

$$V_P = \frac{e}{2x_0}\left[\frac{x_0}{b+x_0} + \frac{4b^3x_0}{(b+x_0)^4-b^4}\right].$$

APPENDIX II

Image Potential for a Prolate Spheroidal Projection

For spheroidal projections the image potential may be expressed in terms of prolate spheroidal harmonics with coordinates ξ, η, θ. Since the problem has axial symmetry, the solution will not contain the longitude angle θ. As in figure 6, the surfaces η constant will represent the spheroidal projections on the plane surface $\xi = 0$.

In this coordinate system the potential of a point charge $-e$ on the axis at $P(1, \eta_0)$ is given by

$$V_1 = -e \sum_{n=0}^{\infty} \frac{2n+1}{b_1} P_n(\xi) P_n(\eta) Q_n(\eta_0) \qquad \eta < \eta_0$$

where $b_1^2 = b^2 - a^2$.

This point charge will induce an equal and opposite charge on the surface of the metal which is at zero potential.

The potential of such a charge distribution will be of the form

$$V_2 = -e \sum_{n=0}^{\infty} [A_n P_n(\xi) + B_n Q_n(\xi)] [A_n' P_n(\eta) + B_n' Q_n(\eta)] \qquad \eta > \eta_1$$

and will be subject to the conditions $V_2 \to 0$ as $\eta \to \infty$ and V_2 finite when $\xi = 1$, whence

$$V_2 = -e \sum_{n=0}^{\infty} A_n'' P_n(\xi) Q_n(\eta).$$

The total potential due to the charge at P will then be

$$V_1 + V_2 = -e \sum_{0}^{\infty} \left(A_n'' P_n(\xi) Q_n(\eta) + \frac{2n+1}{b_1} P_n(\xi) P_n(\eta) Q_n(\eta_0) \right).$$

This potential must be zero on the projection $\eta = \eta_1$ and thus

$$A_n'' = -\frac{2n+1}{b_1} \frac{P_n(\eta_1) Q_n(\eta_0)}{Q_n(\eta_1)}$$

i.e.

$$V_2 = e \sum_{0}^{\infty} \frac{2n+1}{b_1} \frac{P_n(\xi) P_n(\eta_1) Q_n(\eta_0) Q_n(\eta)}{Q_n(\eta_1)}$$

which at the point $P(1, \eta_0)$ becomes

$$V_P = e \sum_{0}^{\infty} \frac{2n+1}{b_1} \frac{P_n(\eta_1) Q_n^2(\eta_0)}{Q_n(\eta_1)}$$

$$= e \sum_{0}^{\infty} \frac{2n+1}{(b^2 - a^2)^{1/2}} \frac{P_n(\eta_1) Q_n^2(\eta_0)}{Q_n(\eta_1)}.$$

Analysis of the Formation Current in Electrolytic Oxidation of Zirconium

By J. J. POLLING and A. CHARLESBY
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

MS. received 22nd September 1953

Abstract. When zirconium is oxidized electrolytically only a portion of the charge passed contributes to the growth of the oxide film produced. It is therefore assumed that the formation current consists of both ionic and electronic components, the former producing film growth, and the latter liberating oxygen. In this paper these components are measured separately.

The ionic current is deduced from the observed weight increase. After correction has been made for dissolution of zirconium and for the trapping of volatile compounds within the film a figure of 0·41 μg cm^{-2} v^{-1} is obtained for the film growth.

The electron current is determined from the rate of evolution of oxygen. If the remainder of the total current is ionic, there should result a weight increase of 0·46 μg cm^{-2} v^{-1}, due to oxidation.

The agreement between these two values, which is within experimental error, shows that the basic assumption, namely that both electron and ion currents flow, is correct.

From the results the current efficiency, rate of formation, dielectric constant and film thickness are obtained. Under the conditions of the experiment the rate of film growth amounts to 30 Å per volt.

§ 1. Introduction

THIN oxide layers may be formed on certain metals such as aluminium tantalum and zirconium, by oxidizing them anodically in suitable electrolytes. The films formed are insulating and quite different in character from the usual porous films formed, for example, on aluminium in an electrolyte of aqueous sulphuric or oxalic acid. Over a wide range the film thickness δ is proportional to the voltage V_f at which the film is formed at room temperature, and varies but slightly with the formation current i. This indicates that the average electric field F_f across the film is approximately constant during the formation process.

If the film thickness is compared with the charge density passed it is found that (except in the case of aluminium at high current densities) only part of the total charge passed is needed to produce the observed film thickness. It has been assumed therefore that the formation current i consists of both an ionic component i_+ which causes film growth, and an electronic component i_- which flows through the oxide layer without any permanent effect. The ratio of ion current to total current is termed the current efficiency η. The relationships

between ion current, electron current and field have been studied for aluminium and zirconium (Charlesby 1953a, b, c, d).

It appeared desirable to obtain more direct evidence as to the existence of these two forms of current. In the theory proposed by Mott (1947) film growth on aluminium is assumed to arise from the movement of metallic ions, under influence of the field, from the base metal through the oxide film to the surface. In the case of electrolytic oxidation the current is carried by positive metallic ions through the oxide film and by negatively charged hydroxyl ions in the electrolyte. These combine at the oxide surface, to form the metallic oxide and cause film growth. In the case of an electron current passing through the film no such chemical combination is possible, and oxygen is liberated in the form of a gas at the surface of the oxide. The film growth and the liberated oxygen therefore provide means of measuring the ion and electron currents respectively.

The problem is best studied for a metal in which the two forms of current are comparable in amount. Zirconium was chosen since it is believed that for this metal the current efficiency is usually about 50%. The bulk oxide is known to contain considerable and variable amounts of water. The methods used in this paper enable the degree of hydration of these thin films of oxide to be studied. The dissolution of the films during electrolysis is also investigated.

In the first part of the paper the ion current is deduced directly from the weight increase of the oxide layer. In the second part the electron current is deduced from the oxygen liberated. By subtracting this electron current from the total current, the presumed ion current is obtained. The film growth corresponding to this calculated ion current can then be compared with that measured directly.

§ 2. Measurement of Ion Current

The ionic charge passed can be determined from the weight of oxygen absorbed although this is a small quantity and difficult to measure accurately. Other methods which are potentially more sensitive depend on measurements of film thickness by capacity changes and optical effects and involve a knowledge of dielectric constant, refractive index, density and surface area of the oxide film. For zirconium none of these quantities is known accurately, so the direct gravimetric method was used.

2.1. *Procedure*

Specimens of zirconium sheet of 0·127 mm thickness were buffed, degreased in trichlorethylene and acetone, washed in water and alcohol and dried in a stream of warm air. They were weighed, and anodized at constant current in an approximately normal solution of ammonium borate, readings being taken of the voltage every fifteen seconds to verify linear rise of potential with time. The specimens were removed from the electrolyte, washed with redistilled water and alcohol, dried and weighed again. The anodized area was measured. Table 1 gives the weight gain for each specimen, 0·46 μg cm^{-2} v^{-1}.

Corrections must be applied to the mean value of the weight gain per volt, obtained above, owing to: (i) a weight loss due to dissolution of the zirconium (or the oxide) in the electrolyte, and (ii) a weight gain due to the trapping of the volatile chemicals (e.g. water) in the film during the formation.

The weight loss due to dissolution was estimated by chemical analysis of the resulting electrolyte. The experiments were carried out in polythene beakers and the solutions kept in polythene bottles to minimize adsorption of the zirconium on the walls of the vessels. The results varied between 0·04 and 0·08 mg of zirconium per 100 ml. of electrolyte, i.e. about 0·02 μg cm^{-2} v^{-1}, which is almost negligible compared with the weight of the film.

The weight of volatile compounds included in the film was estimated by heating some specimens, previously anodized to various potentials, at 435°C and plotting the initial weight losses, thus finding the maximum weight loss per specimen. This method yielded the following figures for the weight of volatile impurity per unit area per volt: 0·062, 0·088, 0·070, 0·087, 0·060, with a mean value of 0·073 μg cm^{-2} v^{-1}.

When the above corrections were applied, the net weight gain due to anodic oxidation was found, namely 0·41 μg cm^{-2} v^{-1}.

2.2. *Results*

Table 1. Weight Gain per Volt by Gravimetric Method

Run	Weight Gain (μg cm^{-2})	Voltage Rise	Weight Gain per volt (μg cm^{-2} v^{-1})
A	25·9	50	0·52
B	50·0	100	0·50
C	40·6	100	0·41
D	68·0	150	0·45
E	71·9	150	0·48
F	84·0	200	0·42
G	100·5	200	0·50
H	103·0	250	0·41
J	183·0	400	0·46
		Mean value of weight gain per volt	0·46
		Corrected value	0·41

§ 3. Measurement of Electron Current

The estimation of the liberated oxygen has been attempted by Güntherschulze Betz using a displacement method but this failed owing to the solubility of the oxygen in the electrolyte. The technique used in this paper is based on Winkler's method, namely, the oxidation of iodide to iodine and the titration of this with thiosulphate. The apparatus used (see figure, p. 207) is a slight modification of that designed by Young (1951) and Polling for the determination of the current efficiency of oxide films on tantalum.

3.1. *Winkler's Method for Determination of Oxygen Dissolved in Electrolyte*

The apparatus contains 150–170 ml. of electrolyte before electrolysis. After the electrolysis is completed 3 ml. manganous sulphate (480 g $MnSO_4$ per litre) and 3 ml. alkaline potassium iodide (360 g NaOH plus 100 g KI per litre) are added to the electrolyte. A white precipitate of manganous hydroxide is formed immediately:

$$MnSO_4 + 2NaOH = Na_2SO_4 + Mn(OH)_2.$$

This is oxidized by the oxygen dissolved in the electrolyte to manganic hydroxide:

$$2Mn(OH)_2 + O = 2MnO(OH) + H_2O.$$

3 ml. sulphuric acid (diluted 1:1 with water) is added to reduce the manganic hydroxide to manganous sulphate in the presence of potassium iodide, liberating iodine:

$$2MnO(OH) + 3H_2SO_4 + 2KI = I_2 + K_2SO_4 + 2MnSO_4 + 4H_2O.$$

The iodine solution is then titrated against a centinormal solution of sodium thiosulphate:

$$I_2 + 2Na_2S_2O_3 = Na_2S_4O_6 + 2NaI.$$

The normality of the iodine solution is the same as that of the original oxygen solution and supposing the oxygen to have been dissolved completely, the total amount of oxygen liberated can be obtained.

A detailed description of the apparatus and of the experimental procedure is given in the Appendix.

3.2. *Results*

In each of the experiments the formation current was maintained constant so that the total charge passed was always known. From the amount of oxygen liberated as obtained by titration the electron current was determined for various values of the formation current. The electron current was always smaller than the formation current, the difference being ascribed to the presumed ion current. The electron current i_-, ion current i_+ and current efficiency η (ion current/total current) are shown in table 2. The ion current is taken as the difference between formation current (i) and electron current.

Table 2. Analysis of Formation Current Densities

Run	i (mA cm^{-2})	i_- (mA cm^{-2})	i_+ (mA cm^{-2})	η
5	5·9	1·35	4·55	0·77
15	5·48	1·53	3·95	0·72
11	2·57	0·790	1·78	0·69
26	1·81	0·530	1·28	0·71
12	1·77	0·860	0·910	0·60
23	1·52	0·580	0·940	0·62
18	0·995	0·195	0·800	0·80
2	0·754	0·249	0·505	0·67

The total ion charge q_+ passed during formation is shown in table 3, together with the corresponding formation voltage rise. Since 96 450 coulombs of ion charge correspond to a deposit of one gramme equivalent of oxygen (or 30·8 g of ZrO_2) the weight increase of the specimen (and the weight of the film) can be obtained. This is given in column four.

Table 3. Estimation of Weight Gain per Volt from Electron Current

Run	i_+ (mA cm^{-2})	q_+ (coul cm^{-2})	Calc. Weight gain (μg cm^{-2})	Voltage rise	Weight gain per volt (μg cm^{-2} v^{-1})
5	4·55	0·28	23·2	59·2	0·39
26	1·28	0·58	48·0	74·0	0·63
23	0·94	0·55	45·6	95·0	0·48
18	0·80	0·226	18·8	50·4	0·37
2	0·505	0·192	15·9	36·0	0·44
			Mean value of weight gain per volt		0·46

From tables 1 and 3 it can be seen that the weight gain per volt estimated from the electron charge (assuming the difference between this and the total charge to be entirely due to ion charge) agrees with the weight gain per volt obtained by direct weighing. From this it may be inferred that the basic assumption is correct, namely, that the total current passing between anode and cathode does in fact consist entirely of ions and electrons, the former going to build up the film, and the latter to liberate oxygen in the electrolyte.

§ 4. Rate of Formation, Formation Field and Dielectric Constant

The formation rate G is usually defined as the voltage rise per minute divided by the formation current i expressed in mA cm^{-2}, i.e. $G=(dV/dt)/i$. It is related to the formation field F_f as follows:

$$F_f=\frac{dV}{d\delta}=\frac{dV}{rdq_+}=\frac{dV}{ri_+dt}=\frac{1}{r\eta i}\frac{dV}{dt}=\frac{1}{r\eta}G\frac{10^3}{60}$$

where δ is the film thickness, r is the film thickness per unit ion charge per cm^2, q_+ is the ion charge passed per cm^2 and η is the current efficiency. If the density of the oxide film is 5·67 g cm^{-3} and its chemical structure is ZrO_2, $r=5{\cdot}67\times10^{-5}$ and $F_f=3\times10^5\,G/\eta$ v cm^{-1}.

Values of F_f deduced from the observed values of G and η are shown in table 4 for various formation rates. In theory F_f varies with the formation current density, but it has been shown (Charlesby 1953 c) that this variation is small. The mean value of F_f is $3{\cdot}3\times10^6$ v cm^{-1} and corresponds to a film growth of 30 Å v^{-1} during formation.

Table 4. Rate of Formation and Formation Field

Run	i (mA cm^{-2})	η	G (v min^{-1}/mA cm^{-2})	F_f (v cm^{-1})
5	5·9	0·77	9·9	$3{\cdot}85\times10^6$
26	1·81	0·71	6·1	$2{\cdot}58\times10^6$
23	1·52	0·62	7·0	$3{\cdot}39\times10^6$
18	0·995	0·80	9·8	$3{\cdot}66\times10^6$
2	0·754	0·67	6·6	$2{\cdot}95\times10^6$
			Mean value of field	$3{\cdot}3\times10^6$

From this value of F_f and the measured capacity C, the value of the dielectric constant can be obtained:

$$CV_f=\frac{\epsilon V_f}{4\pi\delta}=\frac{\epsilon}{4\pi}F_f(\text{e.s.u.}).$$

Measurements of CV_f for a number of films give values close to 8μF v cm^{-2} or $7{\cdot}2\times10^6$ e.s.u. If $F_f=3{\cdot}3\times10^6$ v cm^{-1}, $\epsilon\sim27$. This compares with a value of roughly 20 quoted by one of the writers (Charlesby 1953 c) and 12·5 deduced by Güntherschulze and Betz on the incorrect assumption that the refractive index μ is related to the dielectric constant ϵ by the formula: $\mu^2=\epsilon$. Young (1951) found a similar discrepancy in the case of tantalum. The optical thickness $\delta\mu$ of the thin oxide layer has been found to be 46 Å v^{-1} by Charlesby (1953 c) and 43 Å v^{-1} by Güntherschulze and Betz (1931). With a value of 30 Å v^{-1} for δ, this gives μ equal to 1·5 which is well below the value for baddeleyite (about 2·2) indicating that the structure of the thin film is different. This is confirmed by electron diffraction which reveals a structure which is largely amorphous.

§ 5. Discussion

Chemical analysis of the electrolytes after formation has shown that the amount of zirconium dissolved during anodization is almost negligible, and the weight losses through heating at 435°C show that the amount of volatile matter in the film is small (although not negligible). This volatile matter is probably water of hydration, but may also contain some of the organic solvents used in the preparation. The argument that these thin films of zirconium oxide freshly formed in aqueous electrolytes contain large amounts of water of hydration like the bulk oxides reported by J. W. Mellor is therefore invalidated.

A possible source of error in the experiments was the lack of knowledge of the true surface area of the specimens. Experiments on tantalum by Waber and co-workers (1951) gave results indicating a ratio of true to apparent area of close on 10, using specimens polished by a technique very similar to that used by the writers. This high ratio is very surprising since one would expect considerable variations in the measurements of, for example, capacity changes per unit apparent area depending on the method of polishing. In other experiments with zirconium, aluminium and tantalum, using different polishing techniques the product CV_f per unit apparent area does not vary by more than about 10%. In any case, if one polishing technique is used the results are comparative and any error in the surface area will not affect the main conclusions of this paper.

The efficiency of measurement of oxygen concentration has been assessed by blank runs using platinum electrodes and no serious error is likely to arise from this cause, above a concentration of 0·1 ml. O_2 per litre (equivalent to 0·07 coulomb), the limit of accuracy of Winkler's method.

§ 6. Conclusions

This paper shows that when a specimen of zirconium is anodized at constant current, the total current passed consists of both ions and electrons, the former of which build up the film while the latter liberate oxygen. Winkler's method has enabled the current efficiency (ion current/total current) to be measured fairly accurately.

From the data collected the formation field and dielectric constant have been evaluated, the mean values being $F_f = 3{\cdot}3 \times 10^6$ v cm^{-1} (corresponding to a film growth of 30 Å v^{-1}), and $\epsilon = 27$.

The amount of zirconium or zirconium oxide dissolved in the ammonium borate electrolyte during formation, was found to be almost negligible, and the amount of volatile chemicals (such as water of hydration) included in the film, though not negligible, was small enough to invalidate the theory that the oxide films, being formed in an aqueous solution, contain great quantities of water.

Acknowledgments

The authors wish to thank Dr. H. M. Finniston and Mr. O. Flint for helpful discussion, and Mr. C. E. Austing for the determination of the zirconium dissolved in the electrolyte. They would also like to express their appreciation to Mr. J. Overend for constructing the glass apparatus shown in the figure. This paper is published by kind permission of the Director, Atomic Energy Research Establishment.

References

Charlesby, A., 1953 a, *Proc. Phys. Soc.* B, **66**, 317; 1953 b, *Ibid.*, **66**, 533; 1953 c, *Acta Metallurg.*, **1**, 340; *Ibid.*, **1**, 348.

Guntherschulze, A., and Betz, H., 1931, *Z. Elektrochem.*, **37**, 726.

Mellor, J. W., *Treatise on Inorganic Chemistry*, Vol. VII, p. 129 (London: Longmans, Green).

Mott, N. F., 1947, *Trans. Faraday Soc.*, **43**, 429.

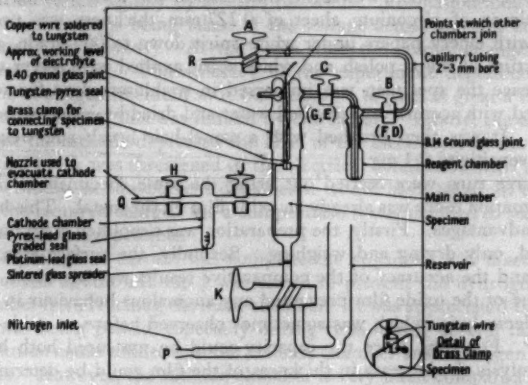
Waber, J. T., Sturdy, G. E., Wise, E. M., and Tipton, C. R., 1951, *Atomic Energy Commission, Unclassified Report* 1355.

Young, L., 1951, *A.E.R.E. Report* M/R 790.

APPENDIX

Description of Apparatus

The apparatus (see figure) is made of Pyrex glass. Access to the specimen in the main chamber is by means of a B.40 ground glass cone and socket. The cathode chamber is kept apart from the main (anode) chamber so that there is little possibility of the products of electrolysis (oxygen and hydrogen) recombining. The cathode is a coil of platinum wire sealed through lead glass which is itself graded to Pyrex.



The apparatus. (For simplicity only one of the three reagent chambers is shown. The others are similar in design and are spaced radially round the main chamber, joining the main chamber and capillary tubing in the diagram at the places indicated.)

The necessary chemicals for the determination—manganous sulphate, alkaline potassium iodide and sulphuric acid—are placed in the three reagent chambers. Access to these chambers is through B.14 cones and sockets. It is clear from the figure that when nitrogen is bubbled through the electrolyte in the main chamber it must also bubble through the reagents on its way out to the atmosphere via stopcock A. Also, by attaching the nitrogen supply to stopcock A, and adjusting the necessary taps, the reagents can easily be blown into the main chamber as required. All stopcocks are lubricated with vaseline grease, and the cones and sockets are fixed in position with wax.

The apparatus is filled by opening all stopcocks except H and K and exhausting it through nozzle R by a water jet pump. The liquid with which it is desired to fill the apparatus is poured into the reservoir and tap K opened. The quickest way of emptying is to attach the water jet pump to outlet Q and to tilt the apparatus. The sintered glass spreader at the bottom of the main chamber is provided to increase the efficiency of the nitrogen sweeping operation, and also to prevent the electrolyte from flowing out when tap K is open to the nitrogen inlet.

Winkler's method assumes that all the oxygen to be estimated is in solution, but in the case of oxygen generated electrolytically this is not so. A large proportion of it exists in the volume above the electrolyte and must be brought into contact with the manganous hydroxide by shaking the apparatus for some time. Preliminary experiments showed that manual shaking was inadequate and an automatic shaker had to be constructed. About two and a half hours of shaking was found to give 90–100% efficiency of estimation of the liberated oxygen.

Experimental Procedure

Before each run the apparatus was completely filled with freshly made chromic acid, left to stand overnight, then washed out with distilled water.

A specimen of zirconium sheet of 0·127 mm thickness was prepared by polishing with emery papers under white spirit down to 000 grade, and finally with a mixture of metal polish and white spirit applied with selvyt cloth. To remove grease the specimen was immersed in trichlorethylene for some time, then washed with acetone and distilled water and dried by immersion in alcohol and ether. It was then brushed with a camel-hair brush and weighed on a balance of sensitivity 0·1 mg.

Successive runs were carried out using the same specimen so that at the start of formation there was already an oxide film on the metal. This had several important advantages. Firstly, the preparation was simplified since no polishing was needed, only drying and weighing. Secondly, the surface area remained the same, and the accuracy of the comparative results was increased. Thirdly, the presence of the oxide film prevented any anomalous behaviour in the initial stages of electrolysis, which was sometimes observed before an oxide layer was deposited.* Fourthly, since the capacity could be measured both before and after electrolysis the increase in thickness of the film could be determined with greater precision.

The specimen was connected to the tungsten wire by means of the small brass clamp shown in the inset of the figure. All the non-zirconium metals were covered with Picien wax to insulate them from the electrolyte, and care was taken to wash the specimen clear of all organic solvents and drying agents since traces of these can result in an iodoform reaction with the iodine and invalidate the titrations.

The two halves of the apparatus were sealed together with hot wax. Taps H and J were closed and the main chamber filled with electrolyte to the

* This behaviour consisted of delayed formation. For several seconds after the current was switched on, no oxide film appeared. At low current densities it was possible to pass current indefinitely without forming an oxide film showing that the current was wholly electronic.

level shown. With tap K closed the apparatus was exhausted to boil off most of the gases dissolved in the electrolyte. All the taps were then closed and the residual electrolyte in the reservoir emptied out and the nitrogen supply connected to the inlet P. It was found advisable not to fill the reagent chambers until the nitrogen was bubbling steadily, since the initial flow of nitrogen was apt to cause violent bubbling in the reagent chambers so that the chemicals were blown out of nozzle R.

When the nitrogen was bubbling steadily and the reagent chambers had each been filled with 4 ml. of the appropriate solution (1 ml. usually remained in each chamber after the rest had been blown out) the apparatus was left for about five hours. Tests made by titrating the resulting solution, without passage of current, showed that this time could not be reduced substantially.

At the end of this period it was assumed on the basis of previous tests that the apparatus was free of oxygen. All the taps were turned off, the vacuum pump was connected to nozzle Q and the cathode chamber evacuated. Tap H was then closed and J opened. The electrolyte almost filled the cathode chamber but left a small bubble of entrapped air. This procedure was repeated three or four times to ensure that the cathode chamber was free of air.

The apparatus was then connected to a circuit in which the voltage across the cell could be varied rapidly by a series resistance. Formation took place at constant current, readings being taken of the voltage across the cell at regular intervals (usually 15 seconds). At the same time the colour changes of the specimen were observed. The accuracy of the experiment depended largely on the volume of oxygen liberated so it was desirable to pass as great a charge as possible. Unfortunately, this was limited by two factors. Firstly, the evolution of hydrogen at the cathode depressed the level of the electrolyte in the cathode chamber until a break was threatened in the electrical connection with the main chamber. Secondly, at low current densities it would have taken an excessive time to pass the required charge. As a compromise it was decided to pass about ten coulombs whenever possible.

When the passage of current was complete the apparatus was disconnected from the electrical circuit and tap J was closed to prevent the hydrogen in the cathode chamber from reaching the main chamber. It was appreciated that while the current was being passed, especially during runs lasting several hours, the oxygen in solution could diffuse into the cathode chamber and hydrogen into the main chamber. However, the cathode chamber is very small compared with the main chamber, the connecting tube is narrow, and rates of diffusion are low, so any error introduced was likely to be negligible. This is supported by the preliminary test run which showed that over 95% of the liberated oxygen could be estimated after prolonged shaking.

The nitrogen supply was connected to nozzle R and tap A turned so that the nitrogen could flow through tube b. This ensured that atmospheric oxygen which had gathered in the tube R and in the tube leading to the nitrogen supply was displaced. When it was deemed that all oxygen had been removed, tap A was turned to admit nitrogen to the capillary tube a. Taps B and C were then opened so that the manganous sulphate was forced into the main chamber. Immediately afterwards the alkaline potassium iodide was added in the same fashion. A white precipitate of manganous hydroxide was formed. All taps were closed, the nitrogen supply was disconnected and the apparatus was shaken

for two and a half hours. The sulphuric acid was then blown into the main chamber to dissolve the precipitate and liberate iodine. Tap K was opened to allow the iodine solution to flow into the reservoir. The first few millilitres were used to wash a clean 25 ml. pipette and the washings were rejected into a measuring cylinder. Samples of the iodine solution were transferred in 25 ml. volumes to a conical flask and titrated against the N/100 thiosulphate contained in a 10 ml. burette, using starch as the indicator. Like the initial washings the remaining iodine solution was poured into the measuring cylinder so that the total volume of the electrolyte could be ascertained. The thiosulphate was itself standardized against a fresh solution of potassium permanganate.

To measure the capacity of the specimen the apparatus was taken apart and the specimen, still connected to the tungsten wire, was washed with distilled water and immersed in a solution of sodium sulphate in an open beaker. A cathode of platinized platinum was also immersed in the solution and the capacity of the oxide film measured with a capacity bridge fed from an a.c. source of 2 volts, and polarized by a d.c. supply of 4·5 volts. Several readings were taken after intervals of some minutes to eliminate initial variations in capacity possibly due to space charge effects.

The specimen was detached from the tungsten wire and the wax dissolved in trichlorethylene. The specimen was washed with acetone and distilled water and again dried with alcohol and ether and weighed.

The colour of the specimen both before and after a run was compared with the colour of a standard specimen (which had been previously anodized in sections at known voltage of 2 minutes) by looking at the specimen normally over a white reflector. This gave a separate estimate of the initial and final voltages of the oxide film.

Finally the geometrical area of the specimen was determined, subtracting from it the area covered by the Picien wax. This non-anodized area was clearly silhouetted by its lack of colour.

The Mechanism and Distribution of Short Period Fading under Conditions of Ionospheric Turbulence

By F. MINOZUMA and H. ENOMOTO
Radio Regulatory Bureau, Ministry of Postal Services in Japan, Tokyo, Japan

MS. received 27th October 1953

Abstract. Fading phenomena are discussed in terms of the variations of optical paths produced by the turbulent ionosphere. A simple model of turbulence in the ionosphere can predict the fading speeds found in practice, and some of the physical constants involved can be deduced from the variation of field intensity with time. The technique can be applied to determine fading speed and the amplitude distributions needed for communication purposes.

§1. Introduction

Continuous measurements of the field intensities for the standard frequency transmissions from JJY (4 Mc/s and 8 Mc/s) have been made at ranges between 90 km and 830 km during one or two weeks, four times a year, since 1948. These measurements have been supplemented by frequency diversity experiments using loop aerials separated by distances between 5 m and 100 m.

Many workers (Briggs *et al.* 1950, Booker *et al.* 1950) have pointed out that the fading of high-frequency waves reflected from the ionosphere is due to an interference phenomenon and originates in the turbulent motions present in the ionosphere. In this paper we attempt to clarify the relations between the distribution and the rate of change of the field intensity and the turbulent phenomena which generate them, using the data obtained from the JJY experiments. We first examine the effect of atmospheric turbulence on the refraction of the waves, then add the contributions from the different paths using the methods developed by Bergman (1946), and hence find the fading characteristics for the wave group.

§2. Turbulence in the Ionosphere

We start by assuming a simple model of atmospheric turbulence in the ionosphere in which the turbulent diffusion coefficient $\lambda^2(z)$ at height z decreases exponentially with height

$$\lambda^2(z) = \lambda_g^2 \, e^{-2z/M}.$$

The mean square perturbation of the atmospheric density $\overline{\Delta\rho^2}$ is given by

$$\overline{\Delta\rho^2} = \lambda^2(z)\rho_0^2(z)$$

or

$$\overline{\Delta\rho^2} = \lambda_g^2 \, e^{-2z/M} \rho_g^2 \, e^{-2z/H} \qquad \ldots\ldots(1)$$

where $\rho_0(z) = \rho_g^2 \, e^{-2z/H}$ is the average atmospheric density at height z.

We also assume that the perturbation is approximately isotropic over distances equal to the atmospheric turbulence scale length l and that l is a constant within one specific ionospheric layer.

When ultra-violet radiation ionizes this turbulent atmosphere, the electron densities produced will fluctuate with position and time. If dS is the radiation absorbed on passing through a section of the atmosphere, thickness dz

$$\left.\begin{aligned} &dS = A\rho(z)S(z)\,dz \\ \text{where}\quad &S(z) = S_0(z) + \Delta S(z) \\ \text{and}\quad &\rho(z) = \rho_0(z) + \Delta\rho(z) \end{aligned}\right\} \qquad \ldots\ldots(2)$$

A being the absorption coefficient per unit volume, $S_0(z)$ the average intensity of radiation at height z and $\Delta S(z)$ the local perturbation of $S(z)$ from $S_0(z)$. We may reasonably expect that our ionization density N is dependent on recombination processes:

$$dN/dt = S - \alpha N^2. \qquad \ldots\ldots(3)$$

Then, solving (3) using perturbation methods, we find that the scale length L for the electron density perturbations is dependent on l and is constant through

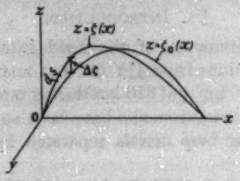


Figure 1. Perturbation of the wave path in a turbulent ionosphere.

the layer. The electron density perturbation ΔN from the average electron density N_0 satisfies

$$\left.\begin{aligned} &\frac{\overline{\Delta N^2}}{N_0^2} = \tfrac{1}{2}\left(\overline{\frac{d\Delta S}{dz}}\right)^2 \Big/ \left(\frac{dS_0}{dz}\right)^2 = K^2 \exp\{(z_M - z)/B\} \\ \text{where}\quad &K^2 = \frac{ML\lambda_g^2}{4H(M+H)};\quad B = \frac{HM}{2(H+M)};\quad z_M = \frac{z_m M}{H+M} \end{aligned}\right\} \qquad \ldots\ldots(4)$$

z_m is the height of maximum electron density. Now the refractive index n in the ionosphere is given by

$$n^2 = 1 - \frac{e^2}{\pi m}\frac{N}{f^2}. \qquad \ldots\ldots(5)$$

Hence the mean perturbation of the square of the refractive index $\overline{\Delta n^2}$ will be

$$\overline{\Delta n^2} \simeq \frac{1 - n_0^2}{4n_0^2}\frac{\overline{\Delta N^2}}{N_0^2} \qquad \ldots\ldots(6)$$

where n_0 is the average refractive index. Since l is constant, the time–space correlation coefficient of Δn between two points separated by time τ and distance r, $\rho_m(\tau, r)$ is independent of the position in the layer.

When a wave is propagated through a medium the wave front is disturbed by the local changes in refractive index. The resultant wave is equivalent to the sum of a large number of independent rays having different phase paths.

The interference between these rays causes the fading phenomena. The optical path length P of a particular ray, shown in figure 1, is

$$\begin{aligned} P = \int n ds &= \int (n_0 + \Delta n)(1 + \zeta'^2 + \eta'^2)^{1/2}\, dx \\ &\simeq \int_0^D n_0 (1 + \zeta_0'^2)^{1/2}\, dx + \int_0^D \Delta n (1 + \zeta_0'^2)^{1/2}\, dx \\ &\quad + \int_0^D \Delta\zeta \frac{\partial n_0}{\partial z} (1 + \zeta_0'^2)^{1/2} dx + \int_0^D n_0 \Delta\zeta \zeta_0' (1 + \zeta_0'^2)^{-1/2} dx \quad \ldots.(7) \end{aligned}$$

in which $n = n_0 + \Delta n$, $z = \zeta(x) = \zeta_0(x) + \Delta\zeta(x)$ (in $x0z$ plane)

and $y = \eta(x) = \eta_0(x) + \Delta\eta(x)$ (in $x0y$ plane)

are the equations of a particular ray, $\Delta\zeta$ is the deviation of the specific path from average path $z = \zeta_0(x)$, $\zeta' = d\zeta/dx$, $\eta' = d\eta/dx$, and D is the distance over the earth's surface.

Since the ionization density at any point varies with time each ray will be phase modulated by $P(t)$. Let us assume that the first and second terms of eqn (7) are predominant, and that $\overline{\Delta\zeta^2}$ is sufficiently large enough compared with L that there is no correlation between individual rays. Then perturbation of the optical path length $\Delta P(t)$ from the average path length $P_0(t)$ follows a normal distribution and

$$\overline{\Delta P(t)^2} \simeq \int_{-\infty}^{\infty} \rho_m(0, s)\, ds \int_0^D \overline{\Delta n^2}(1 + \zeta_0'^2)^{1/2}\, dx = L \int_0^D \overline{\Delta n^2}(1 + \zeta_0'^2)^{1/2}\, dx. \quad \ldots(8)$$

The auto-correlation function of the received power $\rho_Q(\tau)$ due to the combination of the rays (Zadeh 1951, Kiyasu 1951) is

$$\rho_Q(\tau) = \exp\left[-\frac{8\pi^2}{\lambda^2} \int_{-\infty}^{\infty} \{\rho_m(0, s) - \rho_m(\tau, s)\}\, ds \int_0^D \overline{\Delta n^2}(1 + \zeta_0'^2)^{1/2}\, dx \right]. \quad \ldots.(9)$$

If $\overline{\Delta P(t)^2}/\lambda_2^2 \gg 1$, then

$$\rho_Q(\tau) = \exp\left[\frac{4\pi^2\tau^2 L}{\lambda^2} \left\{ \frac{d^2\rho_m(\tau)}{d\tau^2} \right\}_{\tau=0} \int_0^D \overline{\Delta n^2}(1 + \zeta_0'^2)^{1/2}\, dx \right] \quad \ldots.(10)$$

and the fading is deep and its amplitude variation is consistent with a Rayleigh distribution.

§3. Comparison with Experiments

It is convenient to measure the speed of fading using α, the average number of times per second the amplitude passes through the median value (Tanaka and Fujimura 1952). Then

$$\left.\begin{aligned} \alpha &= 2{\cdot}84\sigma \\ \text{where} \qquad \sigma^2 &= -\left\{ \frac{d^2\rho_m(\tau)}{d\tau^2} \right\}_{\tau=0} \frac{\overline{\Delta P(t)^2}}{\lambda^2} \end{aligned}\right\}. \quad \ldots\ldots(11)$$

The fading distribution, which may also be expected to be closely dependent on $\overline{\Delta P(t)^2}$, may be conveniently represented (Kerr 1951) by the first probability distribution of the received power Q.

If the received wave is composed of both steady and variable components and the power ratio of the steady to variable components is m^2, the standard deviation β of Q is given by

$$\beta = \left[\frac{\overline{Q^2} - (\overline{Q})^2}{\overline{Q^2}} \right]^{1/2} = \frac{(1 + 2m^2)^{1/2}}{1 + m^2} \quad \ldots\ldots(12)$$

where β is readily deduced from the experimental observations. Thus m^2 is a convenient parameter to represent the fading distribution. The quantity $v=\overline{\Delta P(t)^2}/LK^2$ can be calculated from the maximum electron density and shape of the layer and expressed in terms of the constants of the atmospheric turbulence, providing the layer is not changing rapidly, for example, as at sunrise or sunset.

The relation between v and α for the resultant signal reflected from the F2 layer during nighttime and daytime are shown in figures 2 and 3 respectively.

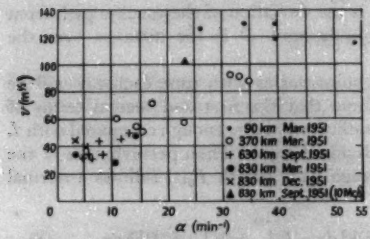


Figure 2. Characteristics of JJY (Tokyo), 4 Mc/s reflected from the F2 layer at night. $z_M=h'_{F2}$, $B=20$ km.

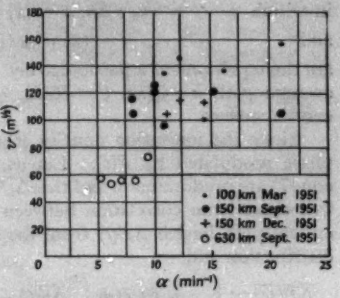


Figure 3. Characteristics of JJY (Tokyo), 8 Mc/s reflected from the F2 layer in the daytime.

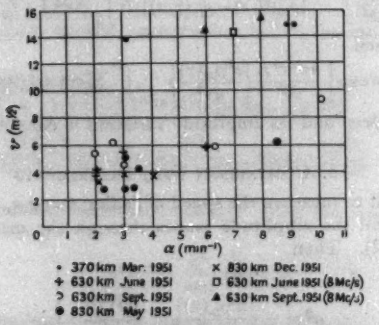


Figure 4. Characteristics of JJY (Tokyo), 4 Mc/s reflected from the E layer in the daytime.

The height z_M has been put equal to the lowest apparent height of reflection in the layer, h'_{F2}, and $B=20$ km was chosen so as to satisfy eqn (10) with the best fit for the experimental results. This corresponds to $H=48$ km and $M=240$ km approximately. Similar data for the single reflection from the E layer, $z_M=110$ km, $B=5$ km are given in figure 4. It appears that the true value of B may be rather larger.

The relation between v^2 and m^2 deduced from the fading distribution for reflections from the F2 layer at night during March 1951 are given in figure 5, again using $z_M = h'_{F2}$, $B = 20$ km.

The curves of best fit in figures 2, 3 and 4 are almost straight lines. They are slightly curved at low fading speeds, which correspond to $\overline{\Delta P(t)^2}/\lambda^2 \ll 1$ and $m^2 \gg 1$ for fading distribution. For these conditions, the optical paths of the component waves are usually not independent, the fading range is small and the distributions are gaussian.

If we define the variation factor V as $\{\overline{(\alpha/v)^2} - \overline{(\alpha/v)}^2\}/\overline{(\alpha/v)}^2$ and compare with the data, the following values are obtained:

Condition	V	Condition	V
F2 layer, night	0·142	F2 layer, daytime	0·063
F2 layer, night ($\alpha > 10$)	0·074	E layer, daytime	0·110

The daily variation of $-LK^2\{d^2\rho_m(\tau)/d\tau^2\}_{\tau=0}$ in the F2 layer for March 1951 is given in figure 6. L is the scale of turbulence, K^2 is the effective intensity of turbulence, and $[-K^2\{d^2\rho_m(\tau)/d\tau^2\}_{\tau=0}]^{1/2}$ is the effective velocity of variation of turbulence in the specific ionosphere. If K^2L, L and the steady speed of wind are determined by space diversity experiments, all the physical quantities which determine the state of turbulence will be obtained. Further work on these problems is proceeding.

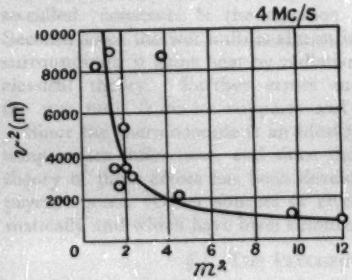


Figure 5. Relation between the turbulence factor v and the fading distribution factor m.

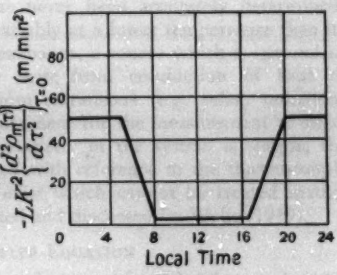


Figure 6. Diurnal variation of effective turbulence.

When interference fading between the various modes of propagation or the magneto-ionic components is present the characteristics of the wave can be calculated provided that the power ratios of the different components are given. For example, the power ratio of the first hop wave to the second hop wave on 8 Mc/s over 630 km shown in figure 3 has been assumed to be 0·84 : 0·16.

We may conclude that the mean square path difference factor $\overline{\Delta P(t)^2}/\lambda^2$ may be a useful index for determining the fading safety factor, but a complete description of the phenomena involves studying K, B, z_M, $\{d^2\rho_m(\tau)/d\tau^2\}_{\tau=0}$ and L at all seasons and during ionospheric storms.

ACKNOWLEDGMENTS

This work has been carried out as part of a project of the Radio Investigation Unit at the Radio Regulatory Bureau and its field stations. The authors wish to thank Mr. W. R. Piggott for his kindness in revising the manuscript for publication, and also thank the courtesy of the Physical Society.

REFERENCES

BERGMAN, P. G., 1946, *Phys. Rev.*, **70**, 486.
BOOKER, H. G., RATCLIFFE, J. A., and SHINN, D. H., 1950, *Phil. Trans. Roy. Soc.*, **242**, 579.
BRIGGS, B. H., PHILLIPS, G. J., and SHINN, D. H., 1950, *Proc. Phys. Soc.*, **63**, 106.
KERR, E. D., 1951, *Propagation of Short Radio Waves, M.I.T.*, **13**, 560.
KIYASU, Z., 1951, *J. Inst. Elec. Comm. Eng. in Japan*, **34**, 952.
TANAKA, K., and FUJIMURA, S., 1952, *26th Joint Meeting of the three Electrical Institutes of Japan.*
ZADEH, L. A., 1951, *Proc. Inst. Radio Engrs., N.Y.*, **39**, 425.

Error and Accuracy in Thermocouple Psychrometry

By J. L. MONTEITH
Department of Meteorology, Imperial College, London

Communicated by B. J. Mason; MS. received 3rd November 1953

Abstract. The paper discusses the main sources of error in the determination of vapour pressure with a thermocouple psychrometer. The predictions of theory and the results of experiment lead to no general agreement on the value of the constant in the classical psychrometer equation. A new derivation of the equation suggests a reduction of the commonly accepted value by a power of the ratio of diffusion coefficients for heat and water vapour. An expression is derived for the radiation error due to the difference of temperature between the wet bulb and its surroundings. Conduction of heat along the thermocouple wires and extraneous radiation produce errors for which an expression is derived for a particular thermocouple model. Finally, wet-bulb temperature errors are related to the corresponding vapour pressure errors.

§ 1. Introduction

The psychrometric determination of humidity, though theoretically simple, is subject in practice to a variety of errors. First, there is only empirical justification for the use of the classical psychrometer equation and the so-called 'constant' of the equation has never been accurately determined. Second, since the wet bulb is almost invariably at a lower temperature than its surroundings, it gains heat by radiative exchange, a source which is ignored in classical theory. Further errors may arise from conduction of heat to the wet bulb from its supports and from extraneous (e.g. solar) radiation.

Since the thermocouple is an ideal instrument for the measurement of small temperature differences, and since the geometry of the system is simple, the theory of these errors has been developed with reference to the thermocouple psychrometer. Other sources of error exist which cannot be treated mathematically and which have been summarized and discussed by Wylie (1949).

§ 2. The Psychrometer Equation

The classical wet-and-dry-bulb theory first postulated by August in 1825 and elaborated by many later workers leads to an equation of the form

$$e_s(T')-e=Ap(T-T') \qquad \ldots\ldots(1)$$

where T is the air temperature (°c), T' the wet-bulb temperature, p the total atmospheric pressure (mm Hg), e the vapour pressure of the air, $e_s(T')$ the vapour pressure of air saturated at T', and A is the psychrometer 'constant'.

In its fully developed form (Whipple 1933) classical theory gives

$$A=\frac{c_p}{L(T')\epsilon}\left(1-\frac{e_s(T')}{p}\right) \qquad \ldots\ldots(1a)$$

where c_p is the specific heat of air at constant pressure, $L(T')$ is the latent heat of condensation of water vapour at T', and ϵ is the ratio of the densities of water vapour and dry air at the same temperature and pressure.

Awbery and Griffiths (1932) measured the dry- and wet-bulb temperatures of homogeneous air samples with mercury-in-glass thermometers aspirated at 3 m sec^{-1} and determined the vapour pressure independently from the dew point or by an absorption method. From these observations, which were made with values of T' between 20°C and 70°C, Whipple found a mean value of $A = 6{\cdot}6 \times 10^{-4}$, with individual values ranging from 2 to 9×10^{-4} but showing no systematic variation with temperature. In the same temperature range, A calculated from eqn (1*a*) varies from 6·38 to $4{\cdot}79 \times 10^{-4}$. On this evidence it seemed possible to neglect the term $e_s(T')/p$, which is principally responsible for the temperature dependence of A.

Müller-Cosna and Maier-Leibnitz (1951) have conducted experiments with a fine-wire thermocouple psychrometer producing air of known humidity by passage over acid of known concentration and temperature. At 200°C they found A to be $6{\cdot}2 \times 10^{-4}$. On theoretical grounds they derived the same value assuming that the gradients of temperature and humidity were constant across the boundary layer and using Polhausen's value of 1/3 for the ratio of boundary layer thicknesses for momentum and heat. This leads to an equation of the form

$$e_s(T') - e = \left(\frac{K}{D}\right)^{2/3} \left(\frac{c_p}{L(T')\epsilon}\right) p(T - T') \qquad \ldots\ldots(2)$$

where K and D are the coefficients of diffusion for heat and water vapour in air. The value of K/D is discussed later.

Wylie (1949) has suggested that, in the absence of radiative effects, the psychrometer constant should depend simply on the ratio of the heat and water vapour transfer coefficients of the wet bulb. Adopting values of these coefficients found by McAdams (1933) and Powell (1940), Wylie gives

$$Ap = \frac{0{\cdot}27 + 0{\cdot}567(Vd)^{0{\cdot}56}}{(Vd)^{0{\cdot}60}} \qquad \ldots\ldots(3)$$

where V (cm sec^{-1}) is the velocity of the air stream normal to a wire of diameter d (cm).

For $Vd = 10$, 100, and 300 cm^2 sec^{-1}, A assumes values of 7·7, 6·4, and $6{\cdot}1 \times 10^{-4}$ ($p = 760$ mm), in apparent agreement with the fact that wet-bulb depressions increase to a maximum with increasing aspiration. It will be shown, however, that this variation can be ascribed entirely to a radiation effect and is therefore not a fundamental property of the wet bulb as (3) suggests. A more serious objection to (3) is that for constant aspiration A increases—and hence the depression decreases—when d is decreased. In practice the depression of a wet bulb increases to a maximum as the dimensions are decreased. A more rigorous application of transfer principles along lines suggested by Jacob (1949) has been found to give an equation in which the 'constant' shows a much weaker dependence on the Reynolds number of flow.

§ 3. A New Derivation of the Psychrometer Equation

Heat transfer by forced convection in a gas is specified by the Nusselt number Nu which, by dimensional analysis, can be expressed as a function of the Reynolds and Prandtl numbers of the flow, i.e.

$$\mathrm{Nu} = \frac{hd}{K\rho c_p} = \phi\left(\frac{Vd}{\nu}, \frac{\nu}{K}\right) \qquad \ldots\ldots(4)$$

where d is a characteristic length, ν the dynamic viscosity of air, ρ the air density, h the heat transfer coefficient and ϕ is a function which must be determined empirically.

The diffusion of small quantities of vapour in air under the action of a concentration field can be similarly represented by a mass transfer coefficient

$$N_M = \phi\left(\frac{Vd}{\nu}, \frac{\nu}{D}\right) \qquad \ldots\ldots(5)$$

by appeal to the principle of similarity of heat and vapour transfer.

The observations of a number of workers for long heated cylinders at right angles to the air stream have been summarized by Jacob (1949) and can be described by the equation

$$\mathrm{Nu} = P\left(\frac{Vd}{\nu}\cdot\frac{\nu}{K}\right)^n = Q\left(\frac{Vd}{\nu}\right)^n \qquad \ldots\ldots(6)$$

where the values of the numerical factors Q and n vary somewhat for different workers but in all cases show a weak dependence on Reynolds number R. The following values are due to Hilpert (1933):

R	1–4	4–40	40–4000
Q	0·891	0·821	0·615
n	0·33	0·39	0·47

When heat and mass transfer occur together we have from (4), (5) and (6)

$$N_M = P\left(\frac{Vd}{\nu}\cdot\frac{\nu}{D}\right)^n \qquad \ldots\ldots(7)$$

i.e.
$$N_M = \mathrm{Nu}\left(\frac{K}{D}\right)^n.$$

The simultaneous use of (4) and (5) implies that the transfers of heat and of vapour occur in non-interacting fields, a condition satisfied by the evaporation of water from a wet-bulb surface. This assumption resembles the 'two-stream' hypothesis of the classical treatment. In the first place, density and temperature differences are small and separation of the air and water vapour by thermal diffusion is therefore negligible. Secondly, since the saturation water-vapour content of air at normal temperatures is of the order of 1%, the surface gradient of vapour does not significantly affect the thermal properties of the air and hence the heat transfer process. This corresponds to the classical assumption that the heat capacity of the vapour can be neglected compared with that of the air.

In the absence of radiative and conductive sources, the heat flux normal to a thermocouple wire of diameter d can be written

$$H = h(T - T') = \mathrm{Nu}\frac{K\rho c_p}{d}(T - T') \qquad \ldots\ldots(8)$$

where T' is now the *observed* wet-bulb temperature.

The water vapour flux is

$$M = N_M\frac{\rho D}{d}\{x_s(T') - x\} \qquad \ldots\ldots(9)$$

where x is the humidity mixing ratio (gm/gm).

In the equilibrium state when T' is constant

$$H = L(T')M. \qquad \ldots\ldots(10)$$

Then combining (8), (9) and (10), and assuming as in the classical derivation that $x_s(T') \simeq \epsilon e_s(T')/p$,

$$e_s(T') - e = A'p(T - T') \qquad \ldots\ldots(11)$$

where

$$A' = \left(\frac{K}{D}\right)^{1-n} \frac{c_p}{L(T')\epsilon}. \qquad \ldots\ldots(11\,a)$$

Equations (1) and (11) differ only by the factor $(K/D)^{1-n}$, which cannot be accurately computed. Montgomery (1947), from independent values of the coefficients, gave $K/D = 0{\cdot}85$. A considerable amount of experimental evidence supports a higher value, and Powell (1940) has found $K/D = 0{\cdot}90$. Combining this with Hilpert's values of n and with $c_p/L\epsilon = 6{\cdot}6 \times 10^{-4}$ we obtain the following values for $A' \times 10^4$ for thermocouple elements of different diameters at various rates of aspiration:

V (cm sec^{-1})	10	50	100	300
d (cm)				
0·01	6·15	6·15	6·19	6·19
0·1	6·19	6·19	6·24	6·24
0·5	6·19	6·24	6·24	6·24

These values are close to the experimental figure of 6·2 found by Müller-Cosna and Maier-Leibnitz.

The additional factor $(K/D)^{1-n}$ arises because the behaviour of the boundary layer of the wet bulb, which is completely ignored in the classical derivation, is introduced through the function ϕ. Since $K < D$ the effect of the boundary layer is to give a slightly greater depression under given conditions than that predicted by (1). The depression decreases slightly with increasing Reynolds number, but in practice such an effect would probably be masked by the opposed and much larger radiation effect discussed in the next section.

The above theory holds only for forced convection when natural convection effects are negligible.

§ 4. Inherent Radiation

In the derivation of (1) and (11) it was assumed that the air flow past the bulb provided the only source of heat. If, however, the surroundings of the wet bulb are at the free-air temperature, the effect of heat gained by radiation may be significant, particularly when the aspiration is slight. Since this effect occurs in all normal psychrometer systems it will be termed the 'inherent radiation' effect.

Following Wylie we may write the flux R of radiant heat normal to the bulb as $h_R(T - T^*)$ where the effective coefficient of radiative heat transfer is $h_R = 4\sigma T^{*3}$, T^* is the observed wet-bulb temperature in °K, ρ is Stefan's constant, and $T - T^* \ll T$. The total transfer of sensible heat can therefore be written

$$H + R = (h + h_R)(T - T^*)$$

whence

$$e_s(T^*) - e = A'p\left(1 + \frac{h_R}{h}\right)(T - T^*). \qquad \ldots\ldots(12)$$

It is convenient to express the effect in terms of the ratio a of the observed depression to the depression calculated from (11). Since $T' \simeq T^*$ we have

$$e_s(T^*) \simeq e_s(T') + \left(\frac{\partial e}{\partial T}\right)_{T^*}(T^* - T'). \qquad \ldots\ldots(13)$$

From (11), (12) and (13)

$$a = \frac{T - T^*}{T - T'} = \left[1 + \frac{r}{1+r}\frac{h_R}{h}\right]^{-1} \qquad \ldots\ldots(14)$$

where

$$r = \frac{A'p}{(\partial e/\partial T)_{T^*}}. \qquad \ldots\ldots(14a)$$

Since h_R is generally much smaller than h, we may write

$$a \simeq 1 - \frac{r}{1+r}\frac{h_R}{h}. \qquad \ldots\ldots(14b)$$

For a 1 cm diameter bulb, a calculated from (14) and (8) is plotted in figure 1 with an experimental curve given by Wylie. Wylie chose T' arbitrarily to make

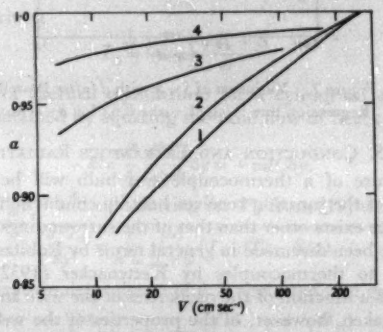


Figure 1. Variation of a with aspiration rate V. 1, d=1 cm : from Wylie's experimental curve, plotted as a/a (300); 2, d=1 cm : from eqn (14), plotted as a/a (300); 3, d=0·122 cm : 18 s.w.g.; 4, d=0·032 cm : 30 s.w.g.

a=1 when V=300 cm sec^{-1}, and if the theoretical values are similarly adjusted agreement is close. The inherent radiation errors of 18 and 30 s.w.g. thermocouple wires are also plotted and show that a approaches unity only slowly as the aspiration rate is increased.

Powell (1936) found a linear relation between a and $\sqrt{d}$ for fine-wire thermocouple, in 'still' air. It was admitted that the psychrometers were mounted in an open-ended wind tunnel in a large room, circumstances not incompatible with an air velocity of several centimetres per second.

From (6), (8) and (14b) we have

$$\log(1-a) = (1-n)\log d - n\log V + \log\left\{\frac{r}{1+r}\frac{h_R v^n}{QK\rho c_p}\right\}.$$

Powell's linear relation with n=0·5 was obtained by drawing a straight line through five of the six observed points. If the same observations are plotted logarithmically (figure 2), and all are given equal weight, the regression line is

$\log(1-a)=0{\cdot}62\ \log d-0{\cdot}46$. This implies that $n=0{\cdot}38$, which compares favourably with Hilpert's value of 0·33 for low Reynolds numbers. Furthermore, this implies that the mean velocity in Powell's tunnel was 5 cm sec^{-1}—a reasonable value consistent with the assumption that (14*b*) is applicable to these observations.

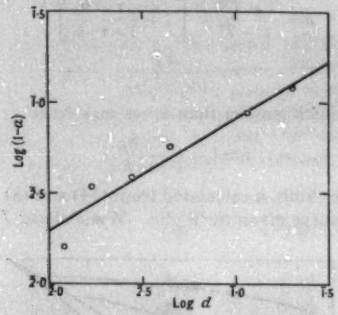


Figure 2. Variation of $1-a$ with d (*after* Powell).
Regression line : $\log(1-a)=0{\cdot}62\ \log d-0{\cdot}46$.

§ 5. Conduction and Extraneous Radiation

The temperature of a thermocouple wet bulb will be greater than that predicted by (12) if the junction receives heat by conduction along the wires or if a radiative source exists other than that of the surroundings at air temperature. These effects have been discussed in general terms by Robitzsch (1932) and with special reference to thermocouples by Kettenacker (1932) who derived an expression for a as a function of the properties of the wire and of the air stream. No account was taken, however, of the properties of the wet wick surrounding the wire—a procedure which cannot be justified *a priori*.

The system discussed here consists of a copper–constantan thermocouple wound with a water-saturated cotton thread and stretched at right angles to the air stream between supports supposed held at air temperature. At 18°c the conductivities of copper and constantan are 0·92 and 0·054 cal cm^{-3} sec^{-1} deg^{-1} respectively, and it is therefore assumed that the heat conducted by the constantan portion of the couple can be neglected if the two portions are of roughly equal length. The wick conductivity is assumed to be 0·0014, the value for water, since the conductivity of dry cotton is very much less. Since, in turn, the conductivity of the wick is much less than that of copper, the radial temperature gradient in the wire can be regarded as negligible compared with that of the wick.

The following additional symbols are required: y = distance along copper wire from support, l = length of wire from support to junction, t = thickness of wick, T_w, T_c = temperatures of wick surface and wire respectively along any radius, k', k'' = thermal conductivities of water and copper, S = gain of radiant heat excluding 'inherent' radiation (cal cm^{-1} sec^{-1}).

If the heat conducted to the wick surface from the wire is written

$$C=\frac{k'(T_c-T_w)}{a\ln(1+t/d)}$$

the heat balance at the surface can be represented by

$$H+R+C+S=L(T_w)M. \qquad \ldots\ldots(15)$$

It is convenient to absorb the inherent radiation in the psychrometer constant by writing $A_R=A'(1+h_R/h)$. Then from (8), (12), etc. and (15)

$$e_s(T_w)-e=A_R p(T-T_w)+\frac{A'p}{h}(C+S). \qquad \ldots\ldots(16)$$

Employing again a modified form of (11) and (13),

$$\left\{A_R p+\left(\frac{\partial e}{\partial T}\right)_{T_w}\right\}(T_w-T')=\frac{A'p}{h}(C+S). \qquad \ldots\ldots(17)$$

Writing now

$$\mu=\frac{C+S}{T_w-T'}=h\left(1+\frac{1}{r(1_w)}\right)+h_R \qquad \ldots\ldots(18a)$$

and

$$\psi=\frac{C}{T_c-T_w}=\frac{k'}{d\ln(1+t/d)} \qquad \ldots\ldots(18b)$$

eqn (17) can be written

$$T_w=\frac{\mu T'+\psi T_c+S}{\mu+\psi}. \qquad \ldots\ldots(19)$$

The equation of thermal equilibrium for a cylindrical element of wire of length δy can be obtained by equating the axial flow of heat to the radial flow at the surface:

$$-k''\pi\frac{d^2}{4}\frac{d^2T_c}{dy^2}\delta y=-\pi dC\delta y. \qquad \ldots\ldots(20)$$

Solving for d^2T_c/dy^2 and substituting for T_w from (19) we find

$$\frac{d^2T_c}{dy^2}=\frac{4}{k''d}\frac{\mu\psi}{\mu+\psi}(T_c-T'-S/\mu). \qquad \ldots\ldots(21)$$

If the slight temperature dependence of μ is neglected, the general solution is

$$T_c=c_1\exp\left\{-\left(\frac{4}{k''d}\frac{\mu\psi}{\mu+\psi}\right)^{1/2}y\right\}+c_2\exp\left\{\left(\frac{4}{k''d}\frac{\mu\psi}{\mu+\psi}\right)^{1/2}y\right\}+T'+\frac{S}{\mu}. \qquad \ldots(22)$$

If the air temperature T is constant or varies only slowly with time it may be assumed that $T_c=T$ at $y=0$, the first boundary condition. The second condition generally adopted in problems of this kind is $c_2=0$. Hence

$$T_c=\left\{(T-T')-\frac{S}{\mu}\right\}\exp\left\{-\left(\frac{4}{k''d}\frac{\mu\psi}{\mu+\psi}\right)^{1/2}y\right\}+T'+\frac{S}{\mu}. \qquad \ldots\ldots(22a)$$

When $y=l$, T_c, the effective temperature of the junction may be written T^*. Then by arrangement of terms

$$a=\frac{T-T^*}{T-T'}=\left\{1-\exp\left[-\left(\frac{4}{k''d}\frac{\mu\psi}{\mu+\psi}\right)^{1/2}l\right]\right\}\left\{1-\frac{S}{(T-T')(\mu+\psi)}\right\}. \qquad \ldots(23)$$

From (18*a*) and (18*b*) we see that μ is related to the heat exchange between the wire and the air stream, i.e. to the aspiration, while ψ is related to the conduction of heat through the wick. In the case $\mu\gg\psi$ (and $S=0$) we see from (19) that $T_w=T'$, i.e. the wick surface attains the 'true' wet-bulb temperature, and from (22*a*) that

$$T^*-T'=(T-T')\exp\left\{-\left(\frac{4\psi}{k''d}\right)^{1/2}l\right\}.$$

The effective wet-bulb temperature is independent of the aspiration rate but may be less than T' since the insulation of the wick may maintain a temperature difference between its surface and the thermojunction beneath it.

In the case $\mu \ll \psi (S=0)$, (19) shows that $T_w < T'$, due to the conduction of heat from the wire to the evaporating surface.

It is obviously important to determine the relative magnitudes of practical values of μ and ψ. The parameter ψ depends only on t and d: values are given in table 1.

Table 1. Values of $\psi = k'/d \ln (1 + t/d)$

d (cm)	0·122	0·032	0·008
s.w.g.	18	30	44
$t/d = 1·0$	0·017	0·063	0·25
$t/d = 0·1$	0·12	0·46	1·8
$t/d = 0·01$	1·2	4·4	17

From (18 *a*) it can be shown that μ is only weakly dependent on wick thickness when this is a small fraction of the wire diameter. (Powell (1936) has demonstrated experimentally that a is only slightly dependent on t in this case.) Values of μ for various V and d have therefore been computed using an arbitrary value of $2t/d = 0·1$ (table 2). Since μ involves also the temperature dependent ratio r, this has been chosen arbitrarily as $\frac{1}{2}$ corresponding to $T^* = 18°$c.

Table 2. Values of $\mu = h\{(1 + 1/r(T_w)\} + h_R$

d (s.w.g.)	18	30	44
$V =$ 1 cm sec^{-1}	0·0012	0·0027	0·0072
$V =$ 10 cm sec^{-1}	0·0025	0·0056	0·015
$V =$ 100 cm sec^{-1}	0·0066	0·015	0·033
$V =$ 1000 cm sec^{-1}	0·019	0·039	0·085

In almost all circumstances, therefore, $\mu \ll \psi$ and (23) can be written

$$a \simeq 1 - \exp\left\{-\left(\frac{4\mu}{k'd}\right)^{1/2} l\right\}. \qquad \ldots\ldots (23\,a)$$

In the next section it will be shown that for accurate work it is desirable that a should be of the order of 0·99. The aspiration rate required to give this value has been calculated as a function of l and d from (23 *a*) and is shown in figure 3. When the wire diameter is fixed by considerations of strength and rigidity, etc., the conduction error can be reduced to zero by ensuring that the depression in a given air stream is independent both of the aspiration rate *and* of the wick length (see Pasquill 1949).

When conduction effects are negligible compared with those of radiation $T^* = T' + S/\mu$ from (22 *a*). In the normal case the wet and dry bulbs are affected by the same extraneous radiation but the effect on the depression depends on the different emissivities of the two bulbs. Since appropriate values of these are difficult to determine, an expression for the combined radiation error, though readily derived, is of little practical value.

§ 6. The Error of Humidity Determinations

It is important to relate psychrometer errors expressed in terms of the fractional error a to the corresponding error in a vapour pressure determination.

Considering first the case in which the error is due to spurious heat sources, the calculated vapour pressure will have an error $\Delta e\,(a)$ where

$$e_s(T^*)-(e+\Delta e(a))=Ap(T-T^*).$$

If the 'true' wet-bulb temperature is defined by (1), then using (13)

$$\Delta e(a)=(1-a)(T-T')\left\{Ap+\left(\frac{\partial e}{\partial T}\right)_{T^*}\right\}. \qquad \ldots\ldots(24)$$

With $a=0{\cdot}99$, $T=20°$C and a relative humidity of 60%, $\Delta e(a)=0{\cdot}08$ mm, i.e. about 0·8% of e. If a drops to 0·90, the error rises to 8%. It is clear that in normal conditions of temperature and humidity (and still more at higher temperatures) a should be of the order of 0·99.

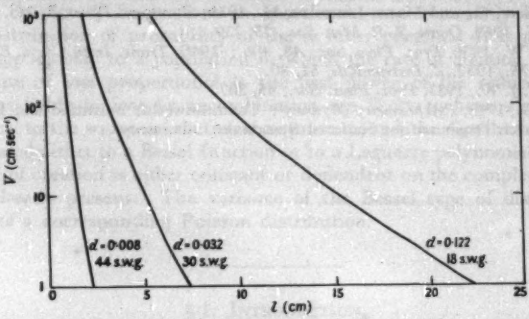


Figure 3. Aspiration rate required to give $a=0{\cdot}99$ for wires of different diameter and length.

When the error introduced by a is small, it may be of the same order of magnitude as that introduced by an inappropriate valuc of the psychrometer constant. If the constant be written $A\pm\Delta A$, then the combined error may be written

$$\Delta e(a,\,A)=\Delta e(a)\mp\Delta Ap(T-T') \qquad \ldots\ldots(24a)$$

if second order terms are ignored. In the above conditions an error in A of only 0·3% is equivalent to $a=0{\cdot}99$.

The error in a determination of A when the vapour pressure is measured independently can now be found by putting $\Delta e(a,\,A)=0$ and combining (24) and (24a) to give

$$\frac{\Delta A}{A}=(1-a)\left\{1+\left(\frac{\partial e_s}{\partial T}\right)_{T^*}\Big/Ap\right\}.$$

In the special case when A is calculated from the classical equation ignoring inherent radiation, $\Delta A/A=h_R/h$. For a bulb of 0·5 cm diameter (e.g. mercury-in-glass thermometer) aspirated at 3 m sec^{-1} in the above conditions, a from (14b) is roughly 0·98 and A is over-estimated by 6%. The accurate determination of the psychrometer constant therefore demands that the wet bulb should be as small as is practical if the aspiration is to be kept within reasonable limits. The fine-wire thermocouple is obviously well suited to such work and indeed to all accurate psychrometry.

Acknowledgments

The author is grateful to Mr. B. J. Mason for his encouragement and advice throughout the work. He would also like to thank the Department of Scientific and Industrial Research for a maintenance allowance.

References

Awbery, J. H., and Griffiths, E., 1932, *Proc. Phys. Soc.*, **44**, 132.
Hilpert, R., 1933, *Forsch. Geb. Ingen.*, **4**, 215.
Jacob, M., 1949, *Heat Transfer* (London : Chapman and Hall).
Kettenacker, L., 1932, *Z. InstrumKde*, **52**, 319.
McAdams, W. H., 1933, *Heat Transmission* (New York : McGraw–Hill).
Montgomery, R. B., 1947, *J. Met.*, **4**, 193.
Müller-Cosna, C., and Maier-Leibnitz, H., 1951, *Z. angew. Phys.*, **3**, 343.
Pasquill, F., 1949, *Quart. J. R. Met. Soc.*, **75**, 239.
Powell, R. W., 1936, *Proc. Phys. Soc.*, **48**, 406 ; 1940, *Trans. Instn Chem. Engrs.*, **18**, 36.
Robitzsch, M., 1932, *Z. InstrumKde*, **51**, 80.
Whipple, F. J. W., 1933, *Proc. Phys. Soc.*, **45**, 307.
Wylie, R. G., 1949, *Psychrometry* (Sydney : Commonwealth Scientific and Industrial Research Organisation, National Standards Laboratory).

Statistics of a Population with Creation and Recombination Dependent on Existing Numbers

By D. A. BELL

Department of Electrical Engineering, University of Birmingham

MS. received 4th November 1953, and in revised form 28th December 1953

Abstract. Studies of fluctuations in gas density are commonly based on a Poisson distribution of probability of the number present. It is pointed out that this applies *only* to a population in which the rate of creation is constant and the rate of loss proportional to the first power of the number present. Populations liable to loss by recombination are likely to have a rate of loss proportional to the square of the number present; and the distribution function is then related either to a Bessel function or to a Laguerre polynomial, according as the rate of creation is either constant or dependent on the complement of the number already present. The variance of the Bessel type of distribution is half that of a corresponding Poisson distribution.

§ 1. Introduction

The classical studies of fluctuations of density in perfect gases (e.g. Chandrasekhar 1943) have been based on the assumption that the number of molecules entering the volume under examination is independent of the number already in it and the number leaving is linearly proportional to the number present. In problems of creation and recombination it is often true that the rate of creation is independent of the existing size of population, but in electrical systems where *pairs* are created, e.g. holes and electrons or positive and negative ions, the rate of recombination may be proportional to the *square* of the number present; and the statistics of the number of particles present are then no longer the same as those of the density fluctuations in a perfect gas. The equilibrium statistics can be found by setting up difference equations for $P_l(t)$, where $P_l(t)$ is the probability of finding l particles present at time t, and setting the time derivative of $P_l(t)$ equal to zero. This is carried out in the next section for rates of creation and destruction constant, proportional to l and proportional to l^2.

§ 2. The Equilibrium Distributions

Feller has pointed out (Feller 1949) that in the case of a series of random events occurring at uniform average rate, the recurrence formula for the probability of exactly l events in time t is satisfied by the Poisson formula for the distribution of l. But there are other cases to be considered and a general method of this kind has been given, for example, by Fry (1928). In his notation p_l and $\bar{p}_l$ are the rates of creation and destruction when the existing

population is l, and he finds the distribution to be described by

$$P(l) = \Delta_l / \Sigma \Delta_l \qquad \ldots\ldots(1)$$

where

$$\Delta_l = \frac{p_0 p_1 p_2 \cdots p_{l-1}}{\bar{p}_1 \bar{p}_2 \cdots \bar{p}_l} \qquad \ldots\ldots(2)$$

and the summation in the denominator of (1) is to be taken over all possible values of l. (In fact it is satisfactory to take it for the full range from $l=0$ to $l=\infty$ in spite of an intrinsic upper limit on l equal to the total available number of particles, provided $P(l) \to 0$ as $l \to \infty$.)

If we assumed a large population sustained at its mean value by random and independent creation and destruction at equal rates, i.e. $p_l = \bar{p}_l =$ constant, equation (2) would reduce to $\Delta_l = 1$ and $\Sigma \Delta_l$ would be unbounded in the limiting case of the total number in the population tending to infinity, so that no useful distribution function results. Such a system is hardly physically realizable since the constancy of population depends on a precise balance between the independent rates of creation and of destruction. An automatic mechanism for maintaining a stable value of mean population is provided by assuming that the rate of destruction is proportional to the number present while the rate of creation is constant. If x is the rate of creation and xl/m the rate of destruction when the magnitude of population is l, equation (2) becomes :

$$\Delta_l = \frac{x^l}{(x/m)^l l!} = \frac{m^l}{l!}. \qquad \ldots\ldots(3)$$

It follows that

$$\Sigma \Delta_l = e^m \qquad \ldots\ldots(4)$$

and the probability distribution of l follows the well-known Poisson formula

$$P_1(l) = e^{-m} m^l / l! \qquad \ldots\ldots(5)$$

where m is the mean value of the population.

If the rate of creation is constant and the rate of loss is proportional to the square of the number present, $\bar{p}_l = al^2$, formula (2) takes the following form:

$$\Delta_l = \frac{(p/a)^l}{(l!)^2}. \qquad \ldots\ldots(6)$$

If this is written

$$\Delta_l = \frac{[(p/a)^{1/2}]^{2l}}{(l!)^2} \qquad \ldots\ldots(7)$$

it can be recognized as a term in the series for a modified Bessel function of the first kind and zero order:

$$\sum_{l=0}^{\infty} \Delta_l = I_0(z) \qquad \ldots\ldots(8)$$

where $z = 2(p/a)^{1/2}$. It follows from (1) that the distribution function is

$$P_2(l) = \frac{(p/a)^l}{(l!)^2 I_0(z)} \qquad \ldots\ldots(9)$$

which is to be compared with the distribution function for the linear case given in formula (5).

When the population has its root-mean-square value n_0 the rates of creation and destruction must be equal:

$$p = a n_0^2. \qquad \ldots\ldots(10)$$

$P_2(l)$ can then be expressed as a function of the r.m.s. number:

$$P_2(l) = \frac{{n_0}^{2l}}{(l!)^2 I_0(2n_0)}. \qquad \ldots\ldots(11)$$

If the particles are randomly created by excitation from a finite number N of sources, the rate of creation will be proportional to $N-l$. Putting $p=a(N-l)$ and $\bar{p}=bl^2$ in (2) we obtain

$$\Delta_l = \frac{a^l N(N-1)\ldots(N-l+1)}{b^l (l!)^2}$$

$$= \left(\frac{a}{b}\right)^l \frac{N!}{(N-l)!}\frac{1}{(l!)^2} = \left(\frac{a}{b}\right)^l \frac{N!}{l!(N-l)!}\frac{1}{l!}. \qquad \ldots\ldots(12)$$

From the form of (12),

$$\Sigma\, \Delta_l = L_N(-a/b) \qquad \ldots\ldots(13)$$

where L_N is the normalized Laguerre polynomial of order N. The distribution function in this case is therefore

$$P_2(l) = [L_N(-a/b)]^{-1} \binom{N}{l} \left(\frac{a}{b}\right)^l \frac{1}{l!} \qquad \ldots\ldots(14)$$

where $\binom{N}{l}$ stands for a binomial coefficient.

§ 3. The Variances of the Distributions

With constant rate of creation and recombination proportional to the first power of the number in the population, there is an equilibrium level of population with a Poisson distribution of probabilities of individual values about the mean. From the properties of the Poisson distribution it follows immediately that the magnitude of the variance is equal to the value of the mean.

With recombination proportional to the square of the number present the variance σ^2 must be calculated from $\sigma^2=\mu_2-(\bar{l})^2$ and the appropriate distribution as given in formula (11) above.

Note that n_0 is by definition the r.m.s. value, so that $\mu_2={n_0}^2$ and to obtain the variance from the second moment one must have also the mean $\bar{l}$ given by:

$$\bar{l} = \Sigma\, lP(l) = [I_0(2n_0)]^{-1} \Sigma\, l{n_0}^{2l}/(l!)^2. \qquad \ldots\ldots(15)$$

The summation can be obtained in terms of the derivative of I_0 which is equivalent to I_1:

$$\bar{l} = \frac{n_0 I_1(2n_0)}{I_0(2n_0)}. \qquad \ldots\ldots(16)$$

For any value of $2n_0$ greater than 10 it will now be found that $\sigma^2=0{\cdot}5n_0$ so that the variance with square-law recombination is exactly half that with linear recombination.*

It is shown in Appendix II that the mean of the Laguerre distribution is

$$l = xL_N'(x)/L_N(x) \qquad \ldots\ldots(17)$$

where x stands for $-a/b$ and the variance is

$$\sigma^2 = [xL_N'(x)+x^2L_N''(x)]/L_N(x) - x^2L_N''(x)/[L_N(x)]^2. \qquad \ldots\ldots(18)$$

* See Appendix I for proof.

The extent to which this can be approximated depends on the relative magnitudes of N and $\lfloor x \rfloor = a/b$. Since the rates of creation and destruction must be equal when $l=\bar{l}$ it follows from the definition of a and b (immediately preceding equation (12)) that $a/b \simeq (\bar{l})^2/(N-\bar{l})$. If something like half the available particles are in the excited state, then $\bar{l}$, N and x are all of the same order of magnitude. The explicit evaluation of σ^2 as a function of N and $\bar{l}$ for the general case appears too cumbersome to be worth while. For the special case of $l \ll N$ it is unnecessary, since the Bessel distribution is a very good approximation to this case.

Acknowledgments

The author wishes to thank Dr. H. P. Mulholland and Mr. J. M. Layton for pointing out the relevance of the Laguerre polynomial and Bessel function to the two types of distribution.

References

Chandrasekhar, S., 1943, *Rev. Mod. Phys.*, **15**, p. 1.

Feller, W., 1949, *Berkeley Symposium on Mathematical Statistics and Probability* (Los Angeles : University of California Press), p. 403.

Fry, T. C., 1928, *Probability and its Engineering Uses* (New York : Van Nostrand).

Appendix I : Limiting Value of $I_1(x)/I_0(x)$

For large values of x one uses the asymptotic series

$$I_n(x) = \frac{e^x}{(2\pi x)^{1/2}}\left\{1 + \frac{1-4n^2}{1!\,8x} + \ldots\right\}$$

the first power in $1/x$ being sufficient for physical 'population' problems where x is likely to be several powers of 10 at least.

Then

$$\frac{I_1(x)}{I_0(x)} = \frac{1-3/8x}{1+1/8x}$$

and

$$1-\left[\frac{I_1(x)}{I_0(x)}\right]^2 = \frac{(8x+1)^2-(8x-3)^2}{(8x+1)^2} = \frac{64x-8}{64x^2+16x+1}.$$

For large values of x this can be approximated by taking only the leading term in the numerator and in the denominator:

$$1-\left[\frac{I_1(x)}{I_0(x)}\right]^2 \simeq \frac{64x}{64x^2} = \frac{1}{x}. \qquad \ldots\ldots(I.1)$$

From equation (16),

$$\sigma^2 = n_0^2\left\{1-\left[\frac{I_1(2n_0)}{I_0(2n_0)}\right]^2\right\}$$

and with the aid of approximation (I.1) this becomes

$$\sigma^2 \simeq n_0^2/2n_0 = 0{\cdot}5n_0. \qquad \ldots\ldots(I.2)$$

Appendix II: Mean and Variance of Laguerre Distribution

The moments of a probability distribution can be found from its characteristic function

$$\phi_v = \Sigma P(l) e^{jvl}.$$

From equation (14) we have for the Laguerre distribution $P_3(l)$ the characteristic function

$$\phi_v = \left[L_N\left(-\frac{a}{b}\right)\right]^{-1} L_N\left(-\frac{ae^{jv}}{b}\right) \qquad \ldots\ldots(II.1)$$

The first moment about the origin is then

$$\mu_1 = \frac{1}{j}[\phi_v']_{v=0} = -\frac{a}{b} L^{-1} L_N'\left(-\frac{a}{b}\right). \qquad \ldots\ldots(II.2)$$

Now $$L_N'(x) = (N/x)L_N(x) - (N^2/x)L_{N-1}(x)$$

therefore $$\mu_1 = L^{-1}[NL_N(-a/b) - N^2 L_{N-1}(-a/b)]. \qquad \ldots\ldots(II.3)$$

Similarly, $$\mu_2 = \frac{1}{j^2}[\phi_v'']_{v=0} = L^{-1}[xL_N'(x) - x^2 L_N''(x)] \qquad \ldots\ldots(II.4)$$

where $x = -a/b$. The variance follows from (II.2) and (II.4)

$$\sigma^2 = \mu_2 - \mu_1^2$$
$$= L^{-1}[xL_N'(x) + x^2 L_N''(x)] - x^2 L^{-2}[L_N'(x)]^2. \qquad \ldots\ldots(II.5)$$

Applying the recurrence formula for $L'(x)$ twice, it is found that

$$L_N''(x) = \frac{N^2 - N}{x^2} L_N(x) - \frac{2(N^3 - N^2)}{x^2} L_{N-1}(x) + \frac{N^2(N-1)^2}{x^2} L_{N-2}(x). \qquad \ldots\ldots(II.6)$$

Since it cannot be assumed that N and x are of different orders of magnitude, it does not seem possible either to replace $L_N(x)$ by a series approximation or to assess the relative importance of the terms containing L_N, L_{N-1}, L_{N-2}. Slight simplification is possible, since N will always be large, by taking only the highest power of N in each coefficient:

$$L_N''(x) \simeq (N^2/x^2)L_N(x) - 2(N^3/x^2)L_{N-1}(x) + (N^4/x^2)L_{N-2}(x). \qquad \ldots\ldots(II.7)$$

A Shearing Interferometer with Fixed Shear and Its Application to some Problems in the Testing of Astro-Optics

By D. BROWN

Sir Howard Grubb, Parsons & Co., Optical Works, Newcastle-upon-Tyne

Communicated by K. J. Habell; MS. received 12th May 1953, and in amended form 26th November 1953

Abstract. A wave-front shearing interferometer is described consisting of a cemented unit which introduces a fixed shear and tilt between the identical interfering wave fronts. This construction results in an instrument which maintains correct compensation and may be very easily and quickly set up and interchanged with other test equipment. The use of such cemented units to measure on and off axis aberrations under workshop conditions is also described.

§1. Introduction

The intention of this paper is to describe the practical applications of the wave-front shearing interferometer in an optical laboratory. Much interest was roused by Dr. W. J. Bates (1947) when he published his paper describing the shearing interferometer which he has constructed and used at the H. H. Wills Laboratory in Bristol. His paper was followed by that of Dr. R. L. Drew (1951) in which a simplified interferometer was described. After experience with Bates and Drew type instruments the writer formed the opinion that for routine testing work in the laboratory a very small instrument with no variable adjustments would be the most valuable form of the interferometer. While the versatility of the Bates instrument is for some research purposes indispensable, the most important properties for routine work are rapidity of action, transportability, and the ability to retain adjustments over considerable periods. These considerations led to the experimental construction of a completely non-adjustable wave-front shearing interferometer which takes the form of a cemented optical unit introducing a fixed shear and tilt between the interfering wave fronts.

§2. Construction of the Instrument

Figure 1 shows the construction of the instrument, which is optically a Drew type interferometer. The angle α controls the shear, which is equal to 2α, the equal angles β control the tilt introduced between the wave fronts. The mirrors M_1, M_2 have been extended for ease of construction, and the mechanical parts are replaced by glass spacers S_1, S_2 which are cut away to clear the light path.

In order to obtain good fringe visibility the angles β must be equal to within $\lambda/8$ over the slit length used. The plane in which the angles β are measured is defined by the (vertical) slit, and the normal to the dividing plate. Since the angle of spacer S_1 is greater than β with the maximum slope inclined at an angle to this plane, depending on the shear, whilst the angle of spacer S_2 is β with maximum slope in the same plane, a slight rotation of slit relative to interferometer about the

axis of the system under test can be used to obtain precise equality of the angles β. The effect of small departures from the calculated angles in the spacers results in an alteration of the shear, disappearance of the fringes at the edges of wide angle cones, and a departure from correct compensation. To obtain fully compensated fringes of high visibility over cones of f/3, the maximum which the instrument will accommodate, the angles of the spacers must be accurate to about 1 second of arc.

§3. Use of the Instrument

In use four simple adjustments have to be made. First, the correct pair of slit images is selected. Next, the instrument is moved along the axis till these images coincide. Thirdly, the instrument or slit source is rotated about the axis of the system under test to make the fringes visible. The fourth adjustment, which is only needed for white light or unfiltered mercury light, consists of equalizing path length in two arms by a translation perpendicular to the optical axis in the shear direction. In practice these can all be made very quickly, and this fact, combined with the small size ($2\,\text{in.} \times 1\tfrac{1}{2}\,\text{in.} \times 1\tfrac{1}{8}\,\text{in.}$) of the instrument, makes possible a rapid substitution of interferometer for knife-edge or other test equipment. This is valuable as the appearance of the optical system under knife-edge test can be used as a starting point in the reduction of the observed fringe pattern.

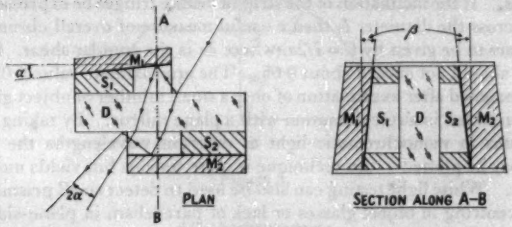


Figure 1.

For small shears Drew has shown that departures from straightness of the central fringe are proportional to slopes on the mirror surface. In some cases it is desirable to convert directly fringe measurement to longitudinal aberration, and for small shears it can easily be shown that $Ar = n\lambda R^2/2SD \tan\phi$, where Ar is the aberration, $n\lambda$ the tilt over the diameter D, R is the radius of curvature, $2S$ the linear shear and ϕ the angle between the shear direction and the line joining the intersections on the interferogram of the central fringe with the zone radius r. In particular, for fixed shear and tilt, $2S/R$ and $n\lambda R/D$ are constants, so the expression becomes $Ar = C \tan\phi$ where C is constant, which is easily measured by calibration on a good spherical mirror. This technique is very useful on highly aspheric wave fronts but it requires measurements made on an interferogram. A visual method with small shear is to traverse the interferometer along the axis until the central fringe intersects a horizontal datum line at the required radius. The longitudinal aberration is then directly measured by the displacement of the interferometer. The above methods do not yield the highest accuracy and, where greater accuracy on highly aspheric surfaces is required, large shear interferometers with zero tilt are useful. The fringe spacing along the line of centres is sufficiently

small to permit a reasonably large number of points to be measured even when taken at the mean focus. A large shear zero tilt interferogram of a highly aspheric wave front is shown in figure 2 (Plate).

For the reduction of interferograms of near spherical wave fronts with moderate or large shear a mechanical technique has been adopted. A mechanical device simulates the action of the interferometer; in the first instance it is supplied with a qualitative shape of the wave front as determined by a knife-edge test, and this is then modified until the observed fringe shape is obtained. This technique, although not elegant, is quick and gives an accuracy to $\lambda/40$ in the reduction of wave fronts of error one wavelength or less.

§4. Chromatic Considerations

For most purposes monochromatic light from a high pressure mercury lamp isolated by means of a gelatine filter has been used, the yellow pair, 5770 Å, 5790 Å, being most useful on account of the high visual and photographic intensity. For certain applications white light may be used, and in this case the region of fringe visibility is restricted to a narrow strip, which for a perfectly achromatic system is parallel to the fringe direction. Imperfect achromatism results in the strip being inclined to the fringe direction owing to the inclination of fringes of the out-of-focus wavelengths. If the inclination of the strip of visible fringes be expressed in terms of fringes across the diameter I, then a useful measure of overall chromatic aberration appears to be given by $C=I/2\alpha$, where 2α is the angular shear. For a good objective C should not exceed about 0·05. The approximate value 0·05 has been arbitrarily assigned after examination of only a small number of object glasses, and applies when tested in autocollimation with a plane mirror. By taking a series of interferograms in monochromatic light at different wavelengths the secondary spectrum may be plotted; this technique is more tedious but yields more specific information. White light testing can also be used to detect small prismatic effects —either decentring in object glasses or lack of parallelism in plane-sided plates. Any prismatic effect parallel to the shear direction will result in a fringe displacement perpendicular to the shear direction and dependent on wavelength. In white light the resulting fringe pattern will no longer have a centro-symmetrical colour sequence.

§5. Coma

Bates (1947) showed that solutions of shearing interferograms for comatic and astigmatic wave fronts are possible, and derived expressions for these cases. However, measurement of coma and astigmatism with the Bates instrument depend on the perfection of its mechanical motions, and in particular the case of astigmatism presents considerable experimental difficulties.

Consider first a comatic wave front which we will represent by $h=ar^3\cos\theta$, where $1>r>-1$ over the aperture. In the $\theta=0$ direction we have $h=ar^3$ as a central cross section. Shearing this cross section by $2S$ and measuring r_1 from the centre of the interference aperture we have

$$h'=a[(r_1+S)^3-(r_1-S)^3]=a(6r_1^2S+2S^3); \qquad \ldots\ldots(1)$$

the term $2S^3a$ represents a fringe displacement perpendicular to the shear direction, the term $6r_1^2Sa$ represents an overall curvature of fringe, proportional to shear. For $S=\frac{1}{4}$ this gives a curvature of the central fringe of $0{\cdot}845a$.

The central cross section of the comatic wave front in $\theta = 90°$ direction is undistorted, so the central fringe in this case remains straight and undisplaced. However, non-central cross sections parallel to it will be distorted. Consider a section XY distant d from centre parallel to the direction $\theta = 90°$. With letters as figure 3 we have $h = ar^3 \cos\theta$, $\cos\theta = d/r$, whence $h = ar^2 d = ad(x^2 + d^2)$. Shearing, and measuring x_1 from centre line of interferogram, we have

$$h' = ad[(x_1 + S)^2 - (x_1 - S)^2] = 4adx_1 S. \qquad \ldots\ldots(2)$$

Putting $S = \frac{1}{4}$ as before we have at $d = \frac{1}{2}$ a slope of $0{\cdot}616a$ across the interference aperture. This slope will have opposite sign when $d = -\frac{1}{2}$, so a total taper of fringe system of $1{\cdot}23a$ will occur.

Referring the comatic wave front to the nearest possible sphere, the maximum error is $\pm a/3$ approximately, so for comatic errors totally $0{\cdot}67a$ (summing + and − errors of $a/3$) sheared by $\frac{1}{4}$ diameter, fringe distortion in the interferogram is $0{\cdot}85a$ and $1{\cdot}2a$ approximately or 1·2 and 1·8 times the coma present when shear is parallel and perpendicular to the flare direction. Figures 4 and 5 (Plate) show a comatic wave front tested with shear respectively approximately perpendicular and parallel to the flare direction.

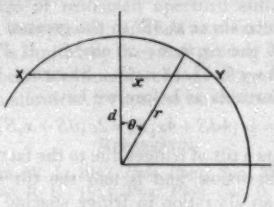


Figure 3.

§6. Astigmatism

Measurement of astigmatism is theoretically quite simple. If the interferometer is placed in focus with the shear direction parallel to the direction of greatest (or least) curvature of the wave front, and then rotated through 90° about the wave-front axis, it will now be out of focus by a distance equal to the separation of the astigmatic focal lines, and this distance can be measured directly. If measurement in terms of wave-front error is required we can consider the fringe rotation due to the difference of curvature. Representing this by $h = br^2$, $1 > r > -1$ as before, shearing by $2S$, and referring to centre of interferogram, we have

$$h' = b[(r_1 + S)^2 - (r_1 - S)^2] = 4br_1 S. \qquad \ldots\ldots(3)$$

For $S = \frac{1}{4}$ we have at the edge of interference aperture $h' = 0{\cdot}75b$. Total tilt over aperture $= 1{\cdot}5b$.

This test is sufficiently sensitive to assess the flatness of plane surfaces placed at an angle in front of a spherical mirror. The experimental difficulty of rotating the interferometer and slit without movement along the axis was overcome by adopting the arrangement indicated in figure 6. The slit and interferometer were mounted at opposite ends of a cut-away tube which rotated inside another tube carrying a pair of diagonal mirrors D_1 D_2 and the condenser C and lamp-house. This arrangement is self-compensating for movement of the inner tube along the tube axis and,

since the directions of rotation of the slit image and interferometer are different, fringes are obtained only at the 0° and 90° positions where measurements are to be made. When used to test inclined unsilvered planes the plane mirror under test and the dividing plate of the interferometer tend to polarize the light, so that with vertical shear the intensity of the image is much less than with horizontal shear. A quarter-wave plate—if of good optical quality—can be introduced immediately in front of the interferometer to equalize the intensities for greater convenience in photography.

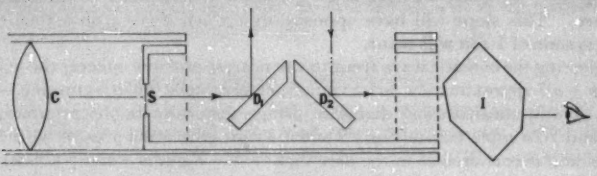


Figure 6.

An alternative method is to shear at 45° to the greatest curvature. Consider an astigmatic wave front of the form $h=cr^2\cos^2\theta$. If PQ$=l$, PR$=d$ (figure 7), $l=r\cos\theta=(d+x)/\sqrt{2}$, $h=cl^2=\frac{1}{2}c(d+x)^2$. Shearing by $2S$ and referring to centre of interference apertures as before we have

$$h'=\tfrac{1}{2}c(4dS+4x_1S)=2c(dS+x_1S). \qquad \ldots\ldots(4)$$

The term $2cSx_1$ represents a tilt of fringes due to the fact that we have not sheared at the mean but at least focus, and is half the tilt encountered in equation (3). The term $2cdS$ is an alteration in fringe spacing which, for $S=\frac{1}{4}$, $d=\frac{1}{2}$, is equal to $c/4$. The displacement measured between $\pm d=\frac{1}{2}$ will then be $c/2$. For a fixed tilt interferometer a single interferogram can be used to determine the

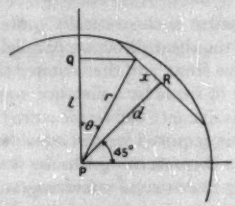


Figure 7.

astigmatism. Higher accuracy can be obtained by taking a second interferogram after rotation of the interferometer through 90°.

In cases where astigmatism and coma are both present with their axes coincident —as for the case of lenses tested off axis—a pair of interferograms taken at the same focus can yield very simply, by measurement of the change of tilt and curvature of the central fringe, a measure of astigmatism and coma. This can be carried out in the presence of spherical aberration, which can be measured on the same interferograms.

§7. Multiple Slits

A further property of fixed shear interferometers is of practical interest. For a shearing interferometer the slit images produced by the two wave fronts are inclined by the shear angle. Considering for simplicity that a path difference of $\lambda/4$ is the greatest permissible for reasonable fringe visibility, we have a restriction of slit width given by $x = \pm\lambda/8\alpha$ where 2α is the shear expressed in radians. This gives a slit width $w = \lambda/4\alpha$. However, we may for fixed shear and monochromatic light put $x = (n\lambda \pm \frac{1}{4}\lambda)/2\alpha$. The permissible source is now a multiple slit, or grating of equal lines and spaces. The grating space is then given by $\lambda/2\alpha$. For good fringe visibility over cones of wide aperture the restriction $n\lambda(1-\cos\psi) < \lambda/4$, where n is half the number of apertures and ψ the semi-angle of the cone, allows $n = 25$ at f/4 , so that a fifty-line multiple slit could be used. In practice multiple slits of up to 50 lines in monochromatic light and up to 20 with unfiltered mercury light have been used with cones of f/4 aperture. A complete mathematical investigation of the use of multiple slits would probably indicate that a considerably greater number of lines could be used without any undue loss of fringe visibility. Even with the number of lines so far used, the increase in intensity of the interference pattern is comparable with that obtainable by autostigmatic testing, and has not the disadvantage of restricted aperture and/or 50% fringe visibility inherent in this method, except when applied to polarizing interferometers. For large shears the method is limited by the small grating space, but to date shears of up to 1/20 radian have been accommodated. The grating space at $\lambda = 5500$ Å for 1/20 radian shear, which allows an f/4 cone to be sheared $d/5$, is $4{\cdot}4 \times 10^{-4}$ inch.

References

Bates, W. J., 1947, *Proc. Phys. Soc.*, **59**, 940.
Drew, R. L., 1951, *Proc. Phys. Soc.* B, **64**, 1005.

The Effect of Compressive and Shearing Forces on the Surface Films present in Metallic Contacts

By M. COCKS

Associated Electrical Industries, Ltd., Research Laboratory, Aldermaston, Berks.

Communicated by W. Hirst; MS. received 28th October 1953

Abstract. Two electrical methods have been used to examine the extent to which the oxide or other natural surface films can prevent intermetallic contact when metal bodies are pressed together. One consisted of measuring the electrical resistance of the contact, the other of measuring the thermoelectric force which arises when there is a temperature gradient across the film; crossed cylinders were used for the contact members. Precautions were taken to minimize vibration and shearing forces when applying normal loads. The contact resistance was often as large as several ohms, showing that penetration of the films was negligible, and in some cases the variation of resistance and thermoelectric force with load was consistent with the assumption that the contact contained a uniform film. Friction forces up to a certain critical value could generally be exerted in the normally loaded contacts before any metallic contact appeared, and in certain cases the films remained intact even during sliding. When the film was penetrated, the area of metallic contact was small at first, but grew as sliding proceeded; there was a simultaneous growth in the force resisting sliding. Microscopic examination showed that the metal surfaces suffered severe damage only after an appreciable distance of sliding (of the order of 0·1 mm).

§1. Introduction

It is generally recognized that metallic friction and the associated surface damage are often largely due to the formation of intermetallic junctions between the sliding surfaces, but relatively little is known about how the junctions form and what determines their size. Intermetallic contact occurs only when the intervening surface films have been penetrated, and the work to be described comprises a study of the conditions under which penetration occurs and of the process by which large-scale breakdown ensues. Electrical contact resistance measurements have been used to determine the extent of metallic contact; the contact members were crossed cylinders. The behaviour of the films has been studied under normal loading, and then under the action of shearing forces progressively increased until large-scale sliding took place. Some of the work has been described in a preliminary note (Cocks 1952).

A number of workers (e.g. Holm 1946, Ross 1941, Umminger 1941) have already measured the resistance between film covered metals under normal loading. In general, the measured resistances have been scattered, and relatively low values of the film resistance have been obtained. It is, however, known that the contact resistance may be very much affected by vibrating the contact members (Bowden and Tabor 1939) and it seemed likely that the scattered nature of the earlier results arose from this cause. In the present work the contacts

were protected from such disturbing influences, and the resistance was usually high enough to prove that the area of metallic contact was minute. With some of the contacts the variation of resistance with load was reproducible and agreed with predictions assuming a uniform film, but usually the resistance was not reproducible and its variation was explained by postulating the existence of tiny punctures in the films. When shearing forces were applied to the contact it was found that the films were not broken up until the force reached a critical value, and with certain metals they remained intact even during sliding. When the film did break up, the area of metallic contact was small at first, but grew as sliding proceeded.

An alternative test for the presence of film is provided by the thermoelectric effect between identical contact members when they are separated by a non-metallic film across which there is a temperature gradient. By the use of this method the interpretation of the contact resistance experiments has been confirmed.

1.1. *Variation of the Contact Resistance with Applied Load*

The resistance due to the constriction in the lines of current flow through a circular contact area of radius a between specimens of the same metal is given by

$$R_c = \rho/2a \qquad \ldots\ldots(1)$$

where ρ is the specific resistance of the contact members (Maxwell 1873). If the contact contains a uniform film of conductivity σ per unit area, the total resistance is given by (Holm 1928)

$$R = \frac{\rho}{2a} + \frac{1}{\pi a^2 \sigma}.$$

It is convenient to rearrange this as follows

$$R = R_c + kR_c^2 \qquad \ldots\ldots(2)$$

where $k = 4/\pi\rho^2\sigma$. When the applied load F is small, the area of contact is determined by elastic deformation of the metal, and

$$a = 1{\cdot}11(r/E)^{1/3}F^{1/3} \qquad \ldots\ldots(3)$$

where r is the radius of the cylinders, and E is their Young's modulus (Hertz 1881).

At large loads extensive plastic deformation occurs and to a good approximation

$$a = (F/\pi P_m)^{1/2} \qquad \ldots\ldots(4)$$

where P_m is a constant known as the flow pressure (Holm 1928).

The variation of R_c with load under elastic or plastic conditions may be obtained by combining eqns (1), (3), and (4). Modified expressions may be derived when the contact members are not of the same metal. If there is a uniform film in the contact, its resistance should vary according to eqn (2) with a suitable value of k.

1.2. *Variation of Thermoelectric Force with Contact Area*

If one member of a contact is heated, the resulting temperature difference t between the two metal–film junctions can cause a thermoelectric force E of magnitude proportional to t even if both members are of the same metal (e.g. Borelius 1919). This temperature difference between the two faces of the film will, in general, be less than the temperature difference between points in the

metal well away from the contact region because of the thermal constriction resistance associated with the contact. This resistance is the thermal analogy of the electrical constriction resistance. Dietrich (1949) has demonstrated that if the film is uniform and T is the temperature difference between the centres of the cylinders, then T/t and, therefore, T/E varies linearly with a. If the film is so thin that t is small compared with T the relationship becomes $E/T \propto 1/a$.

§2. Experimental

2.1. *Apparatus*

The apparatus used for applying forces to the contacts is shown in figure 1. Normal loads were applied by slowly syphoning mercury into the containers A and B, connected through the lever arm C to the carriage D, on which one of the cylinders E was mounted. The carriage was guided by a system of 'sine springs' (Wigan 1949) to ensure that the load was applied normally to the contact area. Movement of the lever arm was opposed by a spring, so that initially the cylinders could be brought together very gently. Shearing forces could subsequently be applied by syphoning mercury into containers F and G connected

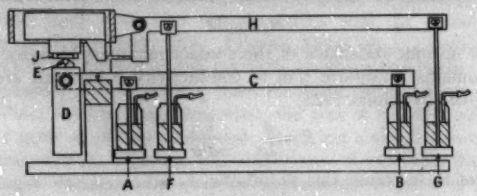


Figure 1. Apparatus for applying normal loads and tangential forces to the contacts. A, B, F, G, loading bottles; C, loading beam; D, carriage for lower cylinder; E, J, crossed cylinders forming the contact; H, tangential loading beam.

through the arm H to the other cylinder J. Relative lateral movements of the cylinders greater than a few tenths of a micron were measured by an optical lever. The apparatus was mounted on a spring suspended table (period about 1 second) resting on a large concrete block mounted on the earth, and independent of the building. If the spring suspension was not used, earth tremors due to passing traffic 30 yards away could cause large fluctuations in contact resistance. The mercury for loading was supplied from movable reservoirs mounted on a separate concrete block, and connected to the apparatus by flexible rubber tubes.

The contact resistance was obtained from voltage–current measurements, the voltage being less than 1 mv to ensure freedom from the heating and coherer effects described by Holm (1946). One of the cylinders could be heated when required and the temperature difference between the centres of the cylinders, usually about 20°c, was measured with a thermocouple.

2.2. *Surfaces of the Contact Members*

Cylinders of a variety of metals were used, and their surfaces were prepared in various ways. Fairly rough surfaces (10–50 microinches centre line average) were produced by grinding or turning. Smoother surfaces (1–5 microinches

centre line average) were prepared by lapping, electrolytic polishing or turning with a diamond plunge tool. The films on the surfaces produced by these techniques were never thick enough to be visible; visible films produced by electrolytic oxidation were used in a few experiments. The electrolytically polished surfaces were rinsed in running tap water and dried; those prepared by other methods were degreased with a solvent.

§3. Results

3.1. *Effect of Applying Normal Loads*

The variation of resistance with applied load was measured at a number of points of contact for each pair of cylinders.

(*a*) *Constriction resistance.* With gold, silver and palladium the measured resistance values agreed closely with those calculated for the metallic constriction resistance, as illustrated for gold in figure 2. The full curve, calculated for increasing loads and drawn on logarithmic scales, consists of a straight line of

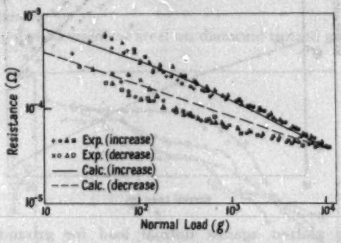


Figure 2. Resistance plotted against normal load for turned gold cylinders. Increase = increasing load at different points of contact. Decrease = diminishing load at different points of contact. Calc = calculated constriction resistance for increasing (or decreasing) load.

slope $-1/3$ given by eqns (1) and (3) at light loads, another of slope $-1/2$ given by eqns (1) and (4) at heavy loads, and a smooth curve joining them. The good agreement with the experimental observations shows that there is no appreciable film resistance, and confirms that the method of calculating the constriction resistance is satisfactory. Contrary to the usual practice, slight vibrations were admitted to the contacts during the measurements on the gold and silver, and this may have helped to break up any films which may have been present; with gold the resistance was slightly higher when the vibrations were suppressed.

When the load is diminished the recovery is purely elastic, and the theoretical resistance–load curve is a single straight line of slope $-1/3$, as shown by the broken line in figure 2. The slow variation of the experimental results as the load began to diminish may have been due to some adhesion of the surfaces. The more rapid increase in resistance at lighter loads could be explained by the breaking of the adhesive junctions by released elastic stresses in the manner described by MacFarlane and Tabor (1950 a). This behaviour was also observed with silver, but with palladium the results followed the theoretical straight line.

(*b*) *Contacts containing uniform films.* With some contacts the experimental resistance variation could be expressed by eqn (2) when a suitable value of k was chosen. As an example, the results for the contact between a ground tungsten cylinder and a turned gold one together with a curve calculated from eqn (2) using a suitably chosen value of k are shown in figure 3. Similar agreement was obtained with a number of different types of contact, and some values of σ calculated from the chosen values of k are given in table 1. The fact that better agreement could have been achieved by using a slightly different value of k for each point of contact suggests that the films were only approximately uniform.

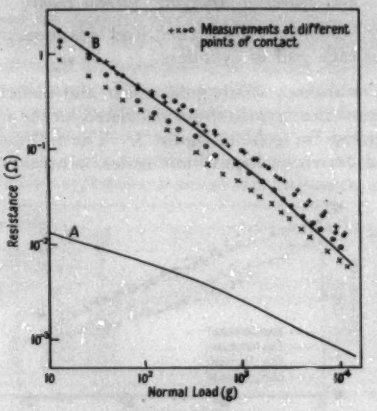


Figure 3. Resistance plotted against normal load for ground tungsten on turned gold. A, calculated constriction resistance $\times 10$; B, calculated, assuming $\sigma = 6{\cdot}5 \times 10^4$ mho cm^{-2}.

Table 1. Values of σ and A_m/A_c (see §3.1 (*d*)) for Contacts containing Uniform Alien Films

Contact Members	σ mho cm^{-2}	A_m/A_c
Turned nickel	$5{\cdot}4 \times 10^7$	3×10^{-3}
Turned nickel on turned gold	$2{\cdot}6 \times 10^6$	4×10^{-4}
Gound tungsten, specimen 1	$1{\cdot}7 \times 10^7$	10^{-2}
Gound tungsten, specimen 2	$2{\cdot}8 \times 10^5$	3×10^{-6}
Ground tungsten (2) on turned gold	$6{\cdot}5 \times 10^4$	3×10^{-7}
Sintered tungsten carbide/cobalt (ground)	$3{\cdot}7 \times 10^6$	4×10^{-3}
Electrolytically polished nickel on gold turned with diamond plunge tool	$2{\cdot}2 \times 10^4$	10^{-9}
Copper after 3 months' exposure to atmosphere	490	0

(*c*) *Contacts containing films which were apparently punctured.* The results for electrolytically polished stainless steel cylinders under increasing load, shown in figure 4, are typical of the majority of the contacts examined. The experimental curves were not reproducible at successive points of contact, and their slopes were generally smaller than was predicted by eqn (2), but they had occasional steep portions. This behaviour is readily explained if most of the conduction

took place through very small regions of intermetallic contact at punctures in the films. The resistance would then be almost independent of the total area of contact, being governed by the size of the metallic areas, which do not increase steadily with the load. When the load was diminished again, sharp drops in resistance often appeared, indicating that the film had suffered slight damage, probably as a result of small shearing forces. The effect was not completely eliminated even when a new apparatus designed with this in mind was used.

Table 2. Typical Values of A_m/A_c at a Normal Load of 1 kg for Contacts whose Films are Punctured

Contact Members	A_m/A_c
Nickel turned with diamond tool	10^{-3}
Lapped nickel	10^{-4}
Electrolytically polished nickel	5×10^{-6}
Electrolytically polished copper	10^{-6}
Electrolytically polished stainless steel	5×10^{-7}
Aluminium alloy (M.G.5) turned with diamond tool	3×10^{-7}
Lapped brass	10^{-6}
Lapped tungsten	10^{-7}
Lapped molybdenum	3×10^{-6}
Electrolytically polished stainless steel on diamond turned gold	3×10^{-6}

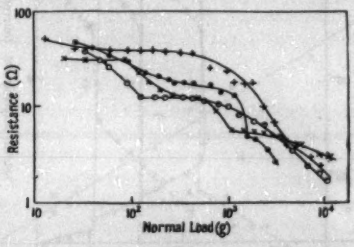


Figure 4. Resistance plotted against normal load for electrolytically polished stainless steel. The curves represent the results obtained at four different points of contact.

(*d*) *Area of metallic contact.* If there is metallic contact through the film, an upper limit to its total area may be determined by assuming that the whole of the conduction is metallic. The metallic contact is represented by a single circular region of area A_m; a smaller total area is obtained if any other shape is assumed, or if it is supposed that there are a number of regions. With contacts, whose resistance variation obeyed eqn (2), calculations of this kind would give too large a value of A_m; with these a smaller value was derived by calculating the area of metallic contact which would just produce a perceptible divergence between eqn (2) and the experimental results. It is convenient to express A_m as a fraction of the load bearing contact area A_c.

A_m/A_c was generally very small but varied considerably with the metal, and for metals whose films were apparently punctured it varied from one point of contact to another. Some typical values are given in tables 1 and 2. The values quoted are by no means the lowest obtained; values of less than 10^{-7} were quite commonly observed, even when there were noticeable punctures in

the films. Moreover, with contact members whose surfaces were so rough that the asperities were considerably deformed, values between 10^{-2} and 10^{-7} were obtained. Furthermore, since A_m is the maximum possible area of metallic contact, the values quoted may be considerably greater than the true areas.

3.2. *Application of Shearing Forces*

(*a*) *Contacts whose films are penetrated during sliding.* Increasing shearing forces were applied to contacts, already under normal load. Figure 5 shows the resulting variation in the resistance and relative displacement for a pair of electrolytically polished stainless steel cylinders. At first the resistance was almost unaffected and the apparent displacement was very small, possibly being due to distortion of the apparatus (in the figure it is exaggerated by the logarithmic scale); when the force reached a critical value the resistance fell sharply and there was a sudden displacement. The critical force was much smaller than

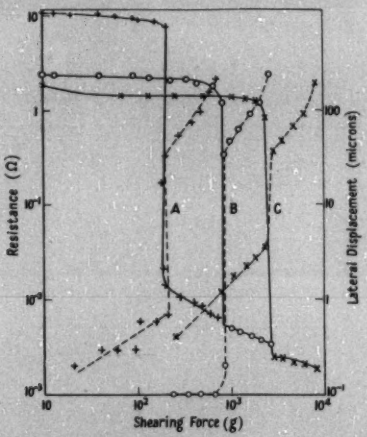


Figure 5. Resistance and displacement for contact between electrolytically polished stainless steel cylinders as shearing force is applied. Normal load: A, 1·1 kg; B, 3·2 kg; C, 10·9 kg.

———— Contact resistance. – – – – – – – – Lateral displacement.

the limiting friction, but the resistance to sliding increased as sliding proceeded. If the shearing force was removed when the critical value was just exceeded and the surfaces separated, microscopic examination showed that the damage was confined to a number of small spots, as shown in figure 6 (*a*) (Plate), but after sliding a few hundred microns it had spread almost all over the area of contact as shown in figure 6 (*b*). The damage probably coincides with the areas of true intermetallic contact. Thus the creation of the first small areas causes the initial drop in resistance, and their growth accounts for the subsequent rise in the force resisting sliding and the slow fall in electrical resistance; this eventually reached a value roughly equal to that calculated from eqn (1) for a film free contact.

This general behaviour was typical of a large number of metals, though there were sometimes minor variations. For example, the initial fall in resistance was sometimes much less sudden, taking place over a two- or three-fold increase in the shearing force. Again, the onset of sliding was sometimes marked, not by a sudden displacement, but merely by an increase in the slope of the measured displacement–force curve.

(*b*) *Contacts whose films resist damage during sliding.* With tungsten, molybdenum and tungsten carbide there was some fall in resistance as the shearing force increased, but the resistance usually remained significantly greater than the value calculated for the film free contact. The sliding which began when a critical force was reached was very rapid, and was not accompanied by any detectable growth in the friction force. Microscopic examination of the surfaces after sliding under small loads showed that the damage was extremely slight, but at loads above about 1 kg small torn patches appeared within the area of contact. With tungsten and tungsten carbide no extensive damage was observed at loads up to 10 kg, the largest load applied, but with molybdenum the films broke down at loads above 3 kg.

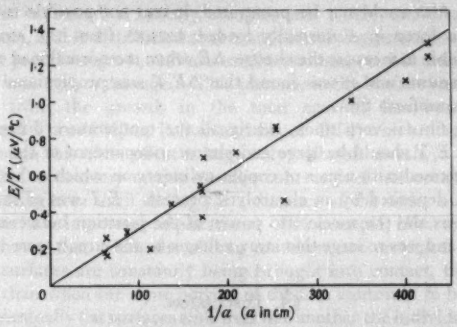


Figure 7. E/T plotted against $1/a$ for turned nickel cylinders.

(*c*) *Contacts between dissimilar metals.* With electrolytically polished stainless steel on diamond-turned gold at loads up to 10 kg and on electrolytically polished copper at loads up to 3 kg, the resistance measurements showed that a large part of the film on the steel remained intact. The surface damage on the steel was confined to a number of thin parallel streaks, and no growth in the friction force was observed once sliding began. With electrolytically polished nickel on gold, however, the film was broken down during sliding at loads as small as a few hundred grammes.

3.3. *Thermoelectric Force Measurements*

With some materials the conclusions drawn from resistance measurements have been confirmed using the thermoelectric effect. The thermoelectric force was difficult to measure because it was small, and stray voltages in the measuring circuit were of the same order of magnitude, despite precautions taken to minimize them. However, if the film is penetrated under the action of a shearing

force, there should be a step in the total measured voltage equal to the thermoelectric force E due to the film. If the film is uniform E/T should be proportional to $1/a$. With turned nickel specimens such a step was observed at the critical shearing force, and figure 7 shows that E/T was in fact proportional to $1/a$.

If under normal loading punctures in the film permit appreciable metallic conduction, the thermoelectric force created at the metal–film interfaces will be at least partly suppressed, and the step in electromotive force when shearing forces are applied should be small. It was found that with electrolytically polished stainless steel or electrolytically polished copper the step was very small, and often undetectable. With electrolytically polished nickel the step was usually observable but was only about 30% of that observed with turned nickel cylinders, even though the electrical resistance was many times greater.

If the film is not penetrated during sliding, the thermoelectric force should not fall to zero; the following procedure was adopted to demonstrate that this was so for molybdenum and tungsten under light loads. The total measured voltage was recorded as sliding proceeded under the light load, and then the load was increased to 10 kg. The voltage fluctuated during sliding, but it never fell as low as the value under the final heavy loading.

Since the film could not be penetrated, it was not possible to separate the thermoelectric force in a normally loaded contact from the strays; it was, however, possible to measure the change, ΔE, when the normal load was increased by various amounts and it was found that $\Delta E/T$ was proportional to $\Delta(1/a)$ as expected for a uniform film.

When the film is very thick, nearly all the temperature difference will be across it, and E/T should be large and almost independent of the contact area. This was confirmed with a pair of copper cylinders on which a thick black CuO film had been deposited by an electrolytic process. E/T was of the same order of magnitude as the thermoelectric power of the junction between copper and CuO in bulk, and was so large that stray voltages in the circuit were unimportant.

§4. Discussion

The resistance measurements have shown that surface films in metallic contacts are seldom penetrated to any appreciable extent under normal loading. Slight penetration often occurs during unloading but this may be due to the difficulty in eliminating shearing forces. In some instances the variation of resistance with load was reproducible and consistent with the assumption of the presence of a uniform film in the contact region. Previous investigators (e.g. Holm 1946, Ross 1941, Umminger 1941) have also obtained results explained by the same assumption. However, their results cannot be used to determine the extent of any penetration of the films, since the resistance values were obtained by taking the lowest or the mean of scattered measurements, and furthermore the film resistances were so low that an appreciable amount of metallic contact could have escaped detection. In the present work the resistance values were generally much higher, and this is most probably due to the precautions taken to minimize vibrations and shearing forces.

The occurrence of intermetallic junctions (which can form only when the intervening films break down) has also been studied by the radioactive tracer method. Rabinowicz and Tabor (1951) have measured the transfer of fragments

from one surface to another resulting from the tearing of these junctions, and they found some transfer even between normally loaded surfaces; they concluded that strong welds were formed. Kerridge (1951) has reported similar experiments covering a greater range of loading; he emphasized that transfer under normal loading is less by several orders of magnitude than that which occurs during sliding. The present work shows that the reason is that the intervening films remained almost intact under normal loading. The contact resistance measurements also suggest that the transfer which did occur was not due to the application of normal loads alone. When shearing forces and vibrations are minimized the area of metallic contact under normal loading often amounts to less than one part in 10^7 of the total contact area even though there may be punctures in the films; the amount of transfer, however, suggested that the metallic junctions occupied an area several orders of magnitude greater than this.

It has often been suggested that sliding causes extensive breakdown of the surface films (e.g. Bowden and Tabor 1945, Finch 1950). However, the present work shows that friction forces up to a critical value can generally be exerted without creating appreciable metallic contact, and with tungsten, molybdenum and tungsten carbide the films often remain intact even during sliding. When metallic contact is created, it initially occupies only a very small fraction of the load bearing area, and is distributed over a number of tiny junctions. As sliding proceeds these junctions grow, and eventually coalesce, so that nearly the whole of the contact area suffers damage. This increase in the surface damage is to be distinguished from the growth in the total area of load-bearing contact (MacFarlane and Tabor 1950 b, Parker and Hatch 1950) which may occur at the same time, though this latter effect is large only with soft metals.

The fact that with many metals metallic contact does not occupy the whole of the load bearing area until sliding for a distance of the order of 0·1 mm had taken place may have important implications. For example, Hirst and Lancaster (1954) have shown that when the sliding members are arranged so that new areas of both surfaces are constantly being brought into contact, the damage is markedly less than when the same portion of the film continues to bear the load. Again, when nominally flat surfaces slide over one another the individual asperities supporting the load are constantly changing, and the distance over which the load is supported on any one asperity is related to its size and shape. If most of the asperities are small, metallic contact may never occupy the whole of the load bearing area even during sustained sliding. Finch (1950) has suggested that the reason why damage is more severe when sliding takes place parallel to the machining lines than when it is perpendicular to them is because the disturbed metal has less chance to escape. This work suggests the alternative explanation that in the former case a given point on the surface remains in contact over a greater sliding distance, so that the welds have a greater chance to grow.

Acknowledgments

The author wishes to thank Dr. W. Hirst for suggesting the problem and for his continued interest and advice. Thanks are also due to Professor R. W. Ditchburn for his interest and helpful suggestions, and to Dr. T. E. Allibone for permission to publish this paper.

References

Borelius, G., 1919, *Ann. Phys., Lpz.*, **60**, 381.

Bowden, F. P., and Tabor, D., 1939, *Proc. Roy. Soc.* A, **169**, 391; 1945, *Chem. Soc. Annual Reports*, **42**, 20.

Cocks, M., 1952, *Nature, Lond.*, **170**, 203.

Dietrich, I., 1949, *Z. angew. Phys.*, **1**, 377.

Finch, G. I., 1950, *Proc. Phys. Soc.* B, **63**, 465.

Hertz, H., 1881, *J. reine angew. Math.*, **92**, 156.

Hirst, W., and Lancaster, J. K., 1954, *Proc. Roy. Soc.* A, in the press.

Holm, R., 1928, *Wiss. Veroff. Siemens-Werk.*, **7**, 217; 1946, *Electric Contacts* (Uppsala : Almquist and Wicksells).

Kerridge, M., 1951, *Proceedings of the Isotopes Techniques Conference*, Oxford, July 1951, **2**, 26 (H.M.S.O. 1952).

MacFarlane, J. S., and Tabor, D., 1950 a, *Proc. Roy. Soc.* A, **202**, 224; 1950 b, *Proc. Roy. Soc.* A, **202**, 244.

Maxwell, J. C., 1873, *A Treatise on Electricity and Magnetism* (Oxford : Clarendon Press), Art. 308.

Parker, R. C., and Hatch, D., 1950, *Proc. Phys. Soc.* B, **63**, 185.

Rabinowicz, E., and Tabor, D., 1951, *Proc. Roy. Soc.* A, **208**, 455.

Ross, A., 1941, *Thesis*, University of Munich.

Umminger, O., 1941, *Thesis*, University of Munich.

Wigan, E. R., 1949, *S.R.D.E. Report* No. 1029.

Mohs's Hardness Scale—A Physical Interpretation

By D. TABOR

Research Group on the Physics and Chemistry of Surfaces,
Department of Physical Chemistry, University of Cambridge

MS. received 24th September 1953

Abstract. The Mohs hardness scale consists of ten minerals arranged in increasing order of hardness such that each mineral will scratch the one on the scale below it, but will not scratch the one above it. Although these minerals are generally considered to be brittle, experiments show that under local indentation and during sliding their behaviour is primarily determined by the plastic properties of the minerals. This suggests that there should be some relation between the Mohs hardness and the indentation hardness of the minerals.

Experiments with metals show that in general a metal surface of indentation hardness H_s will be scratched by a point of hardness H_p if $H_p \geqslant 1{\cdot}2\,H_s$. Thus a scratch hardness scale can be constructed in which each standard has an indentation hardness at least 1·2 times greater than the preceding standard. The scratch hardness number is then proportional to the logarithm of the indentation hardness. Excluding diamond, which is anomalous, this type of relation is found to apply to the Mohs standards, and it is found that each standard is approximately 60% harder than the preceding one.

§ 1. Introduction

It has long been customary for mineralogists and lapidaries to identify or assess the properties of stones or minerals by some type of scratch hardness test. The test which has been most firmly established for more than a century is that due to Mohs (1824) who proposed ten minerals in increasing order of scratch hardness. Each mineral will scratch the one on the scale below it but will not scratch the one above it. At first sight it would appear that such a scale might be so arbitrary as to have no basic physical significance. Mohs himself was aware of this difficulty, and in selecting his standard minerals for the construction of his scratch hardness scale he emphasizes that "the intervals between every two members of the scale be not so disproportionate as either to render its employment more difficult or to hinder it altogether". With these precautions in mind he finally proposed as his ten basic minerals: 1. talc, 2. gypsum, 3. calcite, 4. fluorite, 5. apatite, 6. orthoclase, 7. quartz, 8. topaz, 9. corundum and 10. diamond. Even so Mohs was conscious of the fact that the intervals were not everywhere of the same magnitude, and he points out that the gap between corundum and diamond and, to a lesser extent, between apatite and orthoclase is greater than it should be. Nevertheless, as he concludes in his introductory remarks, "it is very difficult to ascertain the perfect equality of the intervals between the different degrees of hardness, and on that account also it is very difficult to be obtained. Yet all these imperfections are by no means prejudicial to the useful employment of the scale".

Mohs does not describe his criterion for 'perfect equality of the intervals' but in what follows we shall see that there is in fact a sound physical basis upon which equal intervals can be constructed and that the Mohs scale follows it surprisingly well.

§2. The Indentation Hardness of Minerals

In discussing the scratch hardness of minerals it might seem that we are concerned primarily with the behaviour and physical properties of relatively brittle minerals. This, however, is not so. The work of Bridgman (1952) and others has shown that under sufficiently high hydrostatic pressures brittle materials may be prevented from fracturing and, if they are subjected to additional stresses, deformation occurs by the sliding of one row of molecules over another, that is to say plastic deformation takes place. Under localized indentation the stress distribution round the indenter is equivalent to a hydrostatic pressure on which is superposed a shear stress, and in these circumstances brittle fracture is often prevented (King and Tabor 1954). For this reason it is often possible to obtain very satisfactory indentation hardness measurements on relatively brittle materials. Even when some cracking occurs, the cracking often takes place when the indenter is removed. This may be shown by making hardness indentations, for example, in glass and viewing the indentation process through the glass specimen itself. For moderate loads an almost crack-free indentation is formed; on removing the indenter marked cracks are observed; on replacing the indenter many of these cracks close up. It follows that even if the surface shows marked cracking after the indenter is removed, the indentation, if clearly outlined, provides a fairly clear indication of the regions over which plastic flow occurred during the indentation process itself. This is well substantiated by the work of Thibault and Nyquist (1947) who found, for a large variety of brittle materials, that the actual value of the indentation hardness is little affected by the fracturing produced in the surface. Thus the hardness values give a fairly reliable measure of the plastic flow properties of the material under localized contact, when gross brittle fracture is prevented.

In the scratch test itself, where a sharp point or edge is dragged over a flat surface, the conditions at the contact region are similar to those around a static indenter. Here again, although some chipping or fragmentation may occur, friction experiments suggest (King and Tabor 1954) that the deformation process at the rubbing interface is dominated by the plastic deformation of the material rather than by any of its brittle properties. It is thus reasonable to conclude that the scratch hardness behaviour is determined primarily by the plastic properties of the mineral, and we may therefore expect to find some relation between the Mohs hardness and the indentation hardness. This is indeed the case.

Indentation hardness measurements of minerals, in particular of Mohs's ten standards, have been made by Taylor (1949) in England and by Khrushchov (1949) in Russia using a square-based pyramidal indenter like that used in the standard Vickers hardness test. Measurements have been made with the Knoop indenter chiefly by Knoop *et al.* (1939) and by Winchell (1945) in America. The results of these workers are plotted in figure 1. Winchell's results are probably more reliable since the Knoop indenter appears to give less

cracking of brittle materials; in addition he makes it clear that the hardness values depend on the orientation of the indenter as well as on the crystal face.* The general trend is clear. The indentation hardness rises monotonically with increasing steps for each increment of the Mohs scale.

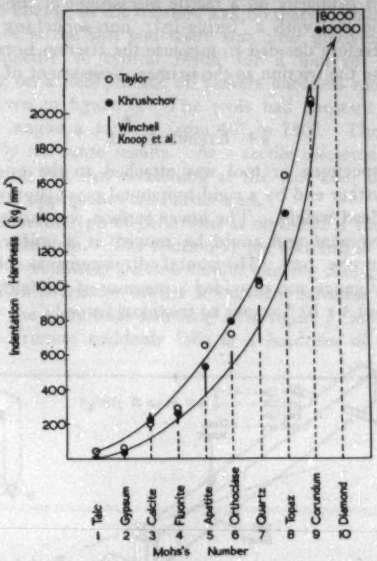


Figure 1. Relation between Mohs's hardness numbers and indentation hardness values.

§3. Assessment of Scratching

Since we are concerned primarily with plastic properties the investigation of the relation between indentation hardness and scratch hardness is greatly simplified by the use of metals instead of minerals. The general procedure was as follows. A strip of metal was prepared, by suitable heat treatment, to be soft at one end and hard at the other, with a fairly uniform increase in indentation hardness along its length. Another metal specimen of intermediate (uniform) hardness had a sharp point ground at one end; the point was pressed with a known load on to the test strip and dragged at a uniform speed over the strip. In the first experiments an attempt was made to assess the occurrence of scratching by the appearance of the track. Over the soft portion of the strip a scratch or groove was clearly visible while at the hard portion scratching had

* Mohs was also aware of the fact that the scratch hardness of minerals often shows a directional variation. The effect, however, does not appear to be very marked for his standard minerals (except possibly diamond) and he does not specify any particular face or direction in establishing his scale. A mineral showing a very large orientation effect is kyanite, which has a scratch hardness on one face of 4 and of 7·5 on another. The corresponding Knoop indentation hardness values according to Winchell are 205 and 1695 kg mm^{-2} respectively, so that both sets of values lie close to the curve of figure 1.

clearly ceased. In some cases a sharp transition was observed, but in most cases the transition was continuous and there was no clear demarcation between the scratched and the unscratched portions of the surface. The reason became clearer when it was realized that in practical scratch testing the occurrence of scratching is based primarily on a tactile assessment of the sliding conditions. Scratching is associated with a 'biting-in'; non-scratching with a 'skidding-over'. It was therefore decided to measure the friction between the point and the strip and to use the friction as the primary assessment of scratching.

§4. Experimental

The pointed specimen or tool was attached to the end of a steel beam supported at its further end by a rigid horizontal pivot (figure 2(*a*)). The load was applied as a dead weight. The lower surface was mounted on a modified lathe traverse movement and could be moved at a uniform speed of about 0·5 cm sec^{-1} under the tool. Horizontal displacements of the beam were measured on a dial gauge and provided a measure of the frictional force. (The sensitivity was about 4×10^{-3} in. per kg frictional force.)

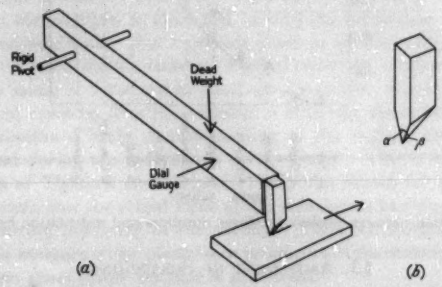


Figure 2. (*a*) Schematic arrangement of apparatus. (*b*) Shape of tool.

Although a conical point is, from the geometric point of view, probably the most satisfactory type to use it is more difficult to prepare. Instead, the point of intersection of three flat ground surfaces was used as the scratching point. This is simple to prepare and resembles more nearly the type of scratching point found on natural minerals or gems. The face angle α and the profile angle β (figure 2(*b*)) could be varied over a wide range.

The lower surface or strip was lightly abraded under water after suitable heat treatment. Indentation hardness measurements were made along the length of the strip with a Vickers indenter. These hardness measurements were made at loads ranging from 2 to 20 kg; it was found that the hardness values were essentially independent of load, showing that the surface was of substantially uniform hardness in depth.

To increase any frictional contrast between scratching and non-scratching, the lower surface was lubricated with oleic acid or a suitable metallic soap.

§ 5. Results

The friction measurements at once revcaled two main types of behaviour. In one type the friction, even at the soft end, was never higher than about $\mu = 0{\cdot}4$, and fell off continuously as the tool approached the harder end of the strip. There was no sharp change in the friction, and no marked change in the appearance of the track. (At the soft end there was clear grooving, but no chip or shaving was produced.) A typical result for a tool of Vickers hardness 940 kg mm^{-2} sliding on a steel surface of Vickers hardness ranging from 350 to 900 kg mm^{-2} is shown in figure 3. The tools had constant profile angles of about 75° and face angles α ranging from 40° to 140°. The different shaped tools gave essentially the same results. As a tactile assessment one would say that there is grooving but not scratching at the soft end, and this gradually diminishes as the tool approaches the harder end.

A completely different type of behaviour is observed if the tool is set up at a different angle to the strip. The friction at the soft end is very high ($\mu \simeq 1$), the motion jerky or intermittent; a fine chip or shaving is often produced and a very marked transition in friction occurs at a critical hardness value of the strip. Typical results (for the same tools and strip as in figure 3) are given in figure 4. It is seen that the friction suddenly falls at a hardness of 800–840 kg mm^{-2}.

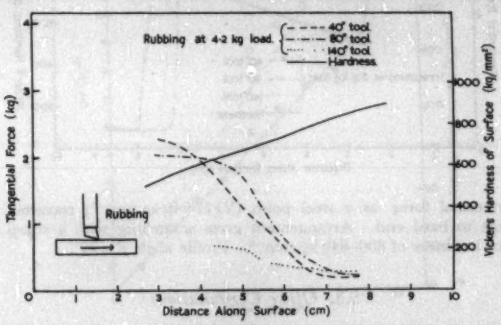


Figure 3. Frictional force as a steel point (VPH 940 kg mm^{-2}) traverses a steel strip from soft to hard end. Arrangement gives grooving and rubbing rather than scratching. Profile angle $\beta = 75°$.

Examination of the track shows a similar transition (figure 5). It is clear that scratching has occurred for hardness values below 800 but not for hardness values above 840 kg mm^{-2}.

5.1. *Effect of Load*

The load was varied from 1·7 kg to 4·2 kg. Essentially the same results were obtained (see figure 5 (Plate)). However, at the smaller load the sharpness of the point was more critical; if the tool was blunt it would not scratch easily. On the other hand repeated tests showed that however carefully the point was prepared the tool of hardness 940 kg mm^{-2} could not scratch a surface of hardness greater than 840 kg mm^{-2}.

5.2. *Effect of Shape of Point*

Experiments were carried out with points of different shape. In the first experiments the profile angle β was kept constant and only the face angle α varied. As the curves in figures 3 and 4 show this has practically no effect on the results. Experiments were then carried out in which the profile angle was varied from 110° to 30°. With 110° it was impossible to obtain scratching. For angles less than 80° the wear of the tool tip was more marked for the tools with smaller profile angles and there was some indication that these tools would not scratch up to so great a hardness. The lowest transition hardness value however was not less than about 750 kg mm^{-2}, and it would seem that there is little dependence of scratch hardness on tool shape. In further work a profile angle of 75° was used throughout.

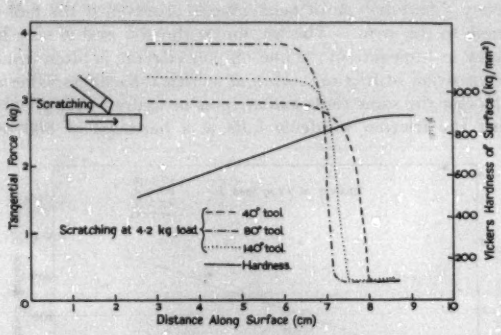


Figure 4. Frictional force as a steel point (VPH 940 kg mm^{-2}) traverses a steel strip from soft to hard end. Arrangement gives scratching with a sharp transition at a surface hardness of 800–840 kg mm^{-2}. Profile angle $\beta = 75°$.

5.3. *Other Combinations*

Systematic measurements were made with other surfaces and tools. The same type of result was obtained. These are summarized in the table, the results in the last line being taken from data quoted by von Weingraber (1952) for a commercial type of ' Ritznadel '.

Material		Vickers hardness (kg mm^{-2})		
Tool	Strip	Tool (H_p)	Transition (H_s)	H_p/H_s
Steel (High speed)	Steel BISRA, CJD2	940	820 (± 20)	1·1–1·2
Steel (High speed annealed)	Steel BISRA, CAD1	600	520 (± 20)	1·1–1·2
Steel (High speed annealed)	Steel BISRA, CAD1	380	300	1·3
β Brass	Hiduminium RR/56	160	130	1·2
Unspecified	All metals	70–80	60	1·3

§ 6. Discussion

These results show that, for a point to scratch a flat surface of Vickers hardness H_s its own hardness H_p must be about 20% greater. The reason for this is not clear. It is not simply related to the difference in yield pressure of a point and a flat surface. If it were so, a point of apex angle 30° would appear to be only about one half as hard as a point with an apex angle of 90° (Tabor 1951, p. 144). In fact the scratch hardness scarcely depends on the shape of the point. For the same reason work hardening of the surface layers by the scratching point cannot be a crucial factor. Nor can the effect be attributed to wear of the tool-tip to a constant final shape. Some Talysurf profilometer records (kindly made by Mr. J. A. Greenwood) show that although there is a gradual decrease in width and depth of the track as the tool moves from the soft to the hard end,

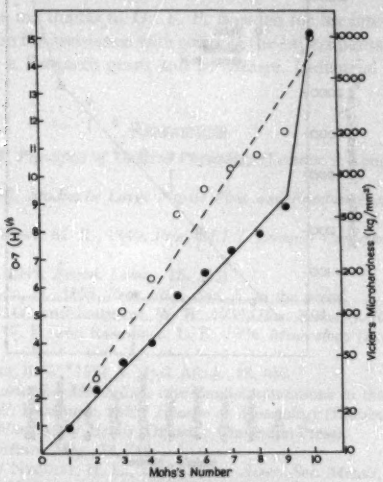


Figure 6. Mohs's hardness number M and indentation hardness H (Krushchov 1949). Full curve shows Krushchov's proposed relation $M=kH^{1/3}$. Dotted curve shows relation $M=k\log H$.

there is no systematic change in the shape of the track up to the stage at which scratching ceases. It would indeed seem that the scratching process, like the cutting process which it closely resembles, is a complex one involving plastic shearing and surface friction; and in spite of the pioneering work of Ernst and Merchant (1940) and the more recent work of Lee and Shaffer (1951), a rigorous solution of the cutting process has not yet been achieved.

Assuming, however, that scratching occurs when $H_p \geqslant 1{\cdot}2H_s$ we may construct a scratch hardness scale in which each standard is at least 1·2 times as hard as the preceding one. If the ratio is exactly 1·2 we should then obtain a scale relating the scratch hardness number M to the indentation hardness H of the form $H=k_1(1{\cdot}2)^M$ or $\log H = M\log(1{\cdot}2)+k_2$. Thus a plot of $\log H$ against M

should give a straight line of slope $\log(1{\cdot}2)$. In practice, of course, it may be desirable to make the gap between each standard somewhat greater to avoid the chance of overlapping, that is, to increase the ratio to something greater than 1·2. The general characteristic, however, would still be a linear relation between $\log H$ and M.

It is interesting to plot the results of figure 1 in this way. Figure 6 shows the results of Khrushchov, who found that for the first nine Mohs standards M is very nearly equal to $kH^{1/3}$. The physical significance of this is obscure. The same values plotted as $\log H$ against M give a fairly straight line with diamond and orthoclase rather high. The results of Winchell and Taylor plotted in figure 7 give a better assessment of the relation between H and M since the range

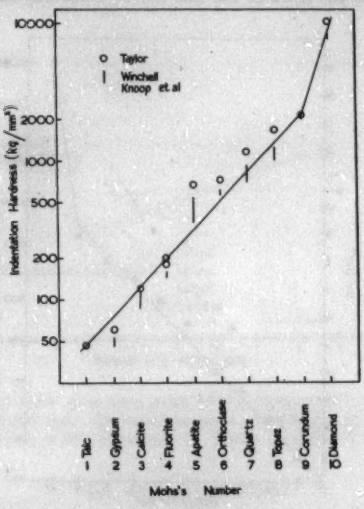


Figure 7. Mohs's hardness number M plotted aginst $\log H$. M is approximately proportional to $\log H$. Each Mohs interval corresponds to an increase in indentation hardness by a factor of about 1·6.

of values of H observed for each mineral is given. It is seen that, excluding diamond, the relation $\log H = nM$ is surprisingly well obeyed. The value of n corresponds to a ratio of hardness between each Mohs standard of about 1·6. It is evident, therefore, that the Mohs hardness scale gives scratch hardness values which correspond to fairly well-defined indentation hardness values, each increment on the Mohs scale corresponding to a 60% increase in indentation hardness. This regularity in behaviour suggests that Mohs did not simply choose 'ten common minerals arranged in order of increasing hardness' (Kraus, Hunt and Ramsdell 1936): it would seem that he experimented with a much larger variety until he had satisfied himself that he had obtained 'equality of the intervals'.

It is interesting to note that in subjective assessments of intensity the least increase of stimulus that is just discernible bears a constant ratio to the original stimulus (Weber–Fechner law): by integration this leads to a linear relation between the logarithm of the stimulus and the measure of the sensation (Bayliss 1918). In estimating equal intervals Mohs presumably used a tactile criterion, and it is probable that under these conditions the same logarithmic law applies.

These results are also of some interest in the abrasion of very hard solids. An abrasive must have a hardness at least 10 to 20% greater than that of the solid if abrasion is to be effective. On this view it is evident that if the same mechanism is valid, diamond abrasives will not be able to abrade the hardest face of diamonds.

Acknowledgments

I wish to express my thanks to Dr. F. P. Bowden for his interest in the work, to K. J. A. Crampton for assistance with some of the experiments, to the Ministry of Supply (Air) for a Research grant and to Messrs. Industrial Distributors for financial support.

References

Bayliss, W. M., 1918, *Principles of General Physiology* (London : Longmans, Green and Co.).

Bridgman, P. W., 1952, *Studies in Large Plastic Flow and Fracture* (London : McGraw–Hill).

Ernst, H., and Merchant, M. E., 1940, *Proc. M.I.T. Summer Conference on Friction and Surface Finish.*

Khrushchov, M. M., 1949, *Zavod. Labor.*, **15**, 243.

King, R. F., and Tabor, D., 1954, *Proc. Roy. Soc.* A, in the press.

Knoop, F., Peters, C. G., and Emerson, W. B., 1939, *Bur. Stand. J. Res., Wash.*, **23**, 39.

Kraus, E. H., Hunt, W. F., and Ramsdell, L. S., 1936, *Mineralogy* (London : McGraw–Hill).

Lee, E. H., and Shaffer, B. W., 1951, *J. Appl. Mech.*, **18**, 406.

Mohs, F., 1824, *Grundriss der Mineralogie* (the English quotations in the paper are from a translation by W. Haidinger, 1825, *Treatise on Mineralogy* (Edinburgh : Constable).

Tabor, D., 1951, *The Hardness of Metals* (Oxford : Clarendon Press).

Taylor, E. W., 1949, *Miner. Mag.*, **28**, 718.

Thibault, N. W., and Nyquist, H. L., 1947, *Trans. Amer. Soc. Metals*, **38**, 271.

von Weingraber, H., 1952, *Technische Härtemessung* (Munich: Carl Hauser).

Winchell, H., 1945, *Amer. Min.*, **30**, 583.

RESEARCH NOTES

Semiconducting Compounds and the Scale of Electronegativities

By C. H. L. GOODMAN

Research Laboratories, General Electric Company, Limited, Wembley, Middlesex

MS. received 15*th October* 1953

In a recent paper, H. Welker (1952) has suggested that electron mobilities in crystals with zinc blende or diamond structures are dependent, among other things, on the ionicity of the bonding, rising to a maximum at some small value of ionicity then falling off as the crystal becomes increasingly heteropolar to reach, eventually, the low values characteristic of essentially ionic crystals. He also suggested that there might be a relationship between ionicity and the energy gap between filled and conduction bands.

Welker restricted attention to the series $C_{diamond}$–Si–Ge–αSn and the compounds between groups III and V of the Periodic Table, such as InSb. He considered that compounds with the zinc blende structure between other groups, such as II and VI, are too greatly heteropolar to show high mobilities, and that their energy gaps are so great that they behave almost as insulators. This may be true of zinc sulphide, but work carried out independently in these Laboratories on the relationship between bond-type and energy gap suggests that this last point is incorrect, and that in some cases reasonably small energy gaps can be obtained with II–VI compounds. Thus CdTe has a gap of approximately 1·5 ev from its absorption spectrum, a value supported by that obtained in these Laboratories from conductivity data (Baker, private communication) of approximately 1·3 ev. It would seem then that Welker's postulate can in fact be extended.

Compound	Type	Δx	ΔE	M. Pt. (degrees)
ZnS	II, VI	1·0	3·7[1] [7]	?
AlSb	III, V	0·6	1·65[2]	1060[2]
CdTe	II, VI	0·6	1·5	1045
InP	III, V	0·5		
InAs	III, V	0·4		942[3] (under pressure)
HgSe	II, VI	0·4		690[4][5]
GaSb	III, V	0·5	0·82[2]	702[2]
HgTe	II, VI	0·2		~670[5]
InSb	III, V	0·5	0·18[6]	523[2]

[1] Piper 1952; [2] Welker 1953; [3] Liu and Peretti 1952; [4] Bloom and Regel 1951; [5] Mokrovskii and Regel 1952; [6] Tanenbaum and Briggs 1953; [7] Wurtzite structure.

A useful guide to the ionicity of the bonding in a binary compound can be obtained from the electronegativity difference Δx of its elements. It must be pointed out, however, that as electronegativity data are of relatively qualitative nature no great accuracy can be claimed in their application, but certain interesting relationships are apparent. The table shows values of Δx for a number of compounds crystallizing with the zinc blende lattice together with their energy gaps ΔE and melting points where known. The Δx values are calculated from the

revised figures given by Haïssinsky (1946), which are more applicable to calculations for solids than those of Pauling. The electronegativity of Hg is taken to be 1·9 and of Sb to be 2·1, the values for divalent and pentavalent atoms respectively. The method of correction for charge effect proposed by Daudel and Daudel (1946) has not been employed as the Δx values are small.

Although electronegativity differences are but approximately related to the ionic component of bonding, it is none the less apparent that there is a correlation between ΔE and Δx. Relatively little information is available about the energy gaps for most compounds with zinc blende lattices, but in a number of cases an approximate value can be predicted if one makes certain simple assumptions : (*a*) that ΔE increases with melting point, as in the series InSb–GaSb–AlSb, (*b*) that compounds with the same melting point and Δx value will have the same value of ΔE, and that increase in Δx will increase ΔE. It is thus possible to make the following predictions: (i) HgTe should have $\Delta E = 0{\cdot}15$–$0{\cdot}4$ ev as its melting point is between those of GaSb and InSb while its Δx is lower. Furthermore, the bonding is stronger than that in αSn ($\Delta E = 0{\cdot}1$ ev) which is unstable above about 20°c. (ii) HgSe should have ΔE in the region of 0·7 ev as its Δx is below those of CdTe and GaSb while its melting point is very close to that of GaSb. (iii) InAs should have ΔE in the region of 0·8–1·0 ev as its Δx lies below those of CdTe and GaSb, while its melting point is intermediate.

That HgSe has been shown to have the high electron mobility of about 7000 $cm^2 v^{-1} sec^{-1}$ (Bloom and Regel 1951) is in agreement with the small ionicity to be expected from electronegativity data, and would seem to indicate that this is not far removed from the optimum value of ionicity postulated by Welker, thus furnishing support for the view that he was incorrect in supposing II–VI compounds to have large ionicities and hence low mobilities. Presumably InSb, with $\Delta x = 0{\cdot}5$ and a mobility of about 25 000 $cm^2 v^{-1} sec^{-1}$ (Welker 1952, Pearson and Tanenbaum 1953), must be closer to this optimum, and hence it is to be expected that HgSe also, with $\Delta x = 0{\cdot}4$, should be close to it and have an extremely high mobility also. Recent Hall effect measurements in these Laboratories on relatively impure specimens of HgSe and HgTe have given electron mobility values in excess of 10 000 $cm^2 v^{-1} sec^{-1}$ (D. McClymont, private communication).

It is to be expected that higher values will be obtained with purer materials. Further work is planned to investigate this point.

The predicted energy gap and high mobility of HgSe suggest that it will prove of particular interest for semiconducting devices such as crystal triodes.

Acknowledgment

The author would like to thank Mr. R. W. Douglas for helpful discussion.

References

Bloom, A. I., and Regel, A. R., 1951, *J. Tech. Phys., U.S.S.R.*, **21**, 316.
Daudel, P., and Daudel, R., 1946, *J. Phys. Radium*, **7**, 12.
Haïssinsky, M., 1946, *J. Phys. Radium*, **7**, 7.
Liu, T. S., and Peretti, E. A., 1952, *Trans. Amer. Soc. Metals*, Preprint No. 40.
Mokrovskii, N., and Regel, A. R., 1952, *J. Tech. Phys., U.S.S.R.*, **22**, 1281.
Pearson, G. L., and Tanenbaum, M., 1953, *Phys. Rev.*, **90**, 153.
Piper, W., 1952, *J. Chem. Phys.*, **20**, 1343.
Tanenbaum, M., and Briggs, H., 1953, *Phys. Rev.*, **91**, 1561.
Welker, H., 1952, *Z. Naturforsch.*, **7a**, 744; 1953, *Ibid.*, **8a**, 248.

The Behaviour of Aluminium Deformed under Alternating Stresses

By N. LOUAT* and M. HATHERLEY†

* Aeronautical Research Laboratories, Department of Supply, Melbourne, Australia
† Defence Research Laboratories, Department of Supply, Melbourne, Australia

MS. received 20th October 1953

A THEORETICAL treatment (Louat, unpublished) has recently been made of the behaviour of face-centred cubic metals under alternating stresses. This work had indicated that, provided the slip lines produced during deformation are single, the active slip planes for the second half cycle of stress should be the same as, or lie less than one micron away from, those operating in the first half cycle. Complete experimental verification of this conclusion would only be possible using the electron microscope technique. However, some useful preliminary information might be expected from an examination with the light microscope, and this note describes some experiments along these lines.

Six tensile specimens, having a gauge length of 2·5 inches and a diameter of 0·357 inch, were machined from a bar of 99·9% pure aluminium and treated by the strain-anneal method of Carpenter and Elam. Grain sizes were produced which varied from the case where the test length of the specimen consisted essentially of a single crystal to the case where the grains were approximately $\frac{1}{8}$ inch in 'diameter'.

The specimens were mechanically and electrolytically polished, lightly etched and subjected to stresses of the order of 10^7 dyn cm^{-2} which were sufficient to produce visible slip lines, although the strains were below the limit of observation (10^{-3}). It was anticipated from the work of Brown (1951) that slip lines produced under these conditions would be single. These lines were examined with a Bausch and Lomb Research Metallograph using oblique lighting, which was found to be more satisfactory than either vertical illumination or the phase contrast technique. A photographic record of the appearance of a typical line was made, the direction of stressing was reversed, and further micrographs of the original slip line were obtained after stressing to various levels.

Typical micrographs obtained with oblique ilumination are shown in figures 1, 2 and 3 (Plate): the magnification ($\times 750$) and obliquity of the lighting is the same in each case. It can be seen that only a small change occurred in the appearance of the slip lines between specimens subjected to $3{\cdot}2 \times 10^7$ dyn cm^{-2} tension (figure 1) and $4{\cdot}0 \times 10^7$ dyn cm^{-2} compression (figure 2). However, when the compressive stress was raised to a value of $6{\cdot}3 \times 10^7$ dyn cm^{-2}, the originally white slip line suddenly became black and vice versa (figure 3) and new slip lines were produced. It is difficult to interpret positively the nature of a change of this type, but it seems to indicate that either the original plane or one close to it had been the site of another large slip movement in the direction opposite to that produced by the initial stressing. The former view appears more probable since reversing the direction of illumination on the specimen as shown in figure 3 produced a facsimile of figure 1. This result would not be expected to follow from the second

explanation since we would then anticipate the formation of either a hollow or a ridge between the two slip planes, and these configurations might be expected to have a similar appearance irrespective of the direction of incidence of the light.

The above phenomena are observed only for the largest stresses, and one must enquire as to the site of deformation in the early part of the second half cycle. No additional slip lines appeared at these low stresses, but there was a slight change in the appearance, under phase contrast, of the original lines with increasing compressive load (cf. figures 1 and 2, and figures 4 and 5). This at least indicates that some change has taken place in the immediate vicinity of the original slip lines. It would seem probable that the operative slip planes are close to the original planes, but that in this case the amount of slip per plane is small. Successive stages in the process might then be represented as in figure 6.

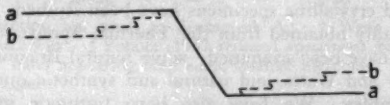


Figure 6. Possible behaviour of a slip step at small reversed strains : a–a, at the end of the first half cycle; b–b early in the second half cycle.

However, in view of the quite abrupt change in appearance of the slip lines under oblique illumination (from black to white and vice versa) it is considered unlikely that all the deformation can be accounted for in this way.

The following tentative conclusion may be drawn from this preliminary work. In an aluminium specimen deformed under alternating stresses slip occurs in the second half cycle on those planes which were operative in the first half cycle, but only at stresses considerably in excess of the maximum stress applied in the first half cycle. This investigation will be continued using the electron microscope.

ACKNOWLEDGMENT

Thanks are due to the Chief Scientist, Department of Supply, Melbourne, Australia, for permission to publish this note.

REFERENCE

BROWN, A. F., 1951, *J. Inst. Met.*, **80**, 115.

On the Formation of Colour Centres in Quartz

By E. W. J. MITCHELL and E. G. S. PAIGE
Physics Research Laboratory, University of Reading

Communicated by R. W. Ditchburn; MS. received 7th October 1953

In this Note we describe results which have been obtained on the formation of colour centres in neutron and x-ray irradiated quartz. The measurements are concerned with the form of the absorption spectrum and the rates at which the various effects occur.

Both fused and crystalline specimens have been studied. The fused quartz was of optical quality obtained from the Thermal Syndicate. Three kinds of crystalline quartz have been examined—some natural Brazilian quartz supplied by Messrs. Hilger and Watts, and natural and synthetic quartz kindly lent by the G.E.C., Wembley. We have also been fortunate in borrowing from the Geology Department of this University a uniformly coloured piece of naturally occurring smoky quartz.

In the range 10 000–2000 Å measurements were made using a Hilger Uvispek spectrophotometer, and below 2300 Å absorption curves were determined for some specimens with a fluorite vacuum spectrograph. All the measurements on crystalline specimens were made with light passing perpendicular to the *c* axis; uniformly coloured crystals were chosen. The reflection losses have been determined so that we are able to calculate the values of the absorption coefficients.

Figure 1 shows the absorption curves of crystalline and fused quartz which have been irradiated with neutrons in the Harwell Pile (BEPO). These are typical of a large number of curves and the features labelled occur regularly.

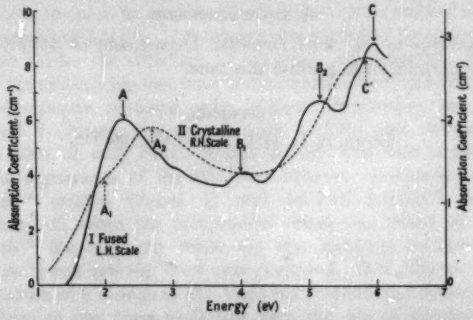


Figure 1. Comparison of absorption spectra of neutron irradiated fused and crystalline quartz.

The irradiation of the amorphous form, in which only short range order is believed to be present, produces curves of similar shape to those obtained with

crystalline quartz, but with differences in detail. The exposure of quartz to x-rays produces effects which are similar to those obtained by neutron irradiation, as will be seen from the table.

Energies (in ev) of Absorption Bands in Irradiated Quartz

Specimen	Irradiation	A_1	A_2	A	B_1	B_2	C
Natural crystalline I	Neutron	2·0	2·65	—	—	—	5·85
Natural crystalline II	Neutron	2·0	2·70	—	—	—	5·95
Natural crystalline II	X-ray	2·0	2·70	—	—	—	5·80
Synthetic	Neutron	*	2·60	—	—	—	5·85
Synthetic	X-ray	*	2·60	—	—	—	5·80
Smoky	?	2·0	2·65	—	—	—	6·0†
Bleached smoky	Neutron	2·0	2·75	—	—	—	5·70
Fused	Neutron	—	—	2·30	4·1	5·2	5·80
Fused‡	X-ray	—	—	2·30	4·1	5·6	5·87

Notes: * Absorption too small to distinguish A_1. † Weak band at 6·0 with a subsidiary feature at 5·4 ev. ‡ Yokota (1952 reduced specimens gave bands at 2·30, 4·1 and 5·6 with a 5·87 band appearing after some thermal bleaching; in oxidized specimens the 2·30 band was absent.

It will also be seen that there are two major variations from the band structure which we have regularly found: the ultra-violet bands found by Yokota (1952) in x-ray coloured fused quartz and the weak C band which we obtained with naturally occurring smoky quartz. The smoky quartz was thermally bleached and after irradiation in the pile for one week gave an absorption spectrum similar to other crystalline quartz.

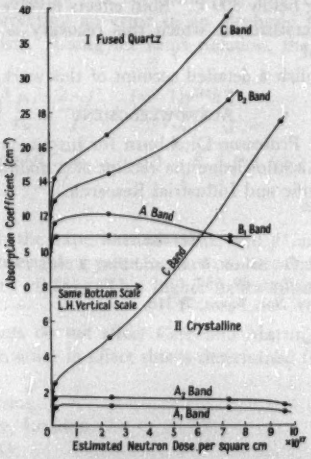


Figure 2. The variation of absorption coefficient with estimated neutron dose.

The variation of absorption coefficient, at the absorption maxima, with the estimated neutron dose is plotted in figure 2. The various bands behave in one of two ways: (i) the ulra-violet bands, C and B_2, increase rapidly with dose,

(ii) the visible bands, A, A_1, A_2 and B_1 decrease slowly after the initial increase. All the specimens have given results of this kind but there are variations between specimens of differing origins in the rates of change of absorption coefficient of a given band. Concerning the second effect, Dr. Berman has informed the authors privately that one of the specimens which he used for his thermal conductivity measurements was colourless after being exposed in the Chalk River pile for one year.

Irradiations with 200 kv x-rays have been carried out in the Royal Berkshire Hospital. The exposure obtained so far, $3{\cdot}5 \times 10^6$ röntgens, gives an absorption of $9{\cdot}4\ cm^{-1}$ in the C band of a natural crystal, which is equivalent to a neutron dose of $0{\cdot}16 \times 10^{17}\ cm^{-2}$ for this type of specimen. Much longer irradiations are necessary to decide whether the slow decrease in visible absorption is also obtained with x-ray coloured specimens.

Measurements are being made on the rates of decay of the absorption bands. Although irradiating with visible light produces no change in absorption, all bands are bleached by irradiating with light in the C band. All the bands are also bleached by heating to temperatures greater than about 250°c, and at 500°c thermal bleaching is complete in a few minutes. The work of Wittels (1953) and of Berman *et al.* (1950), however, has shown that there are other effects associated with neutron bombardment which are more difficult to remove. Wittels found that large increases in the lattice parameters were produced and that normal spacings were obtained only after heating to temperatures of 900°c. Similarly, Berman *et al.* have shown that the increase in thermal resistivity at low temperatures, which is found after irradiation, cannot be removed completely by heating below 500°c. Both effects involve recovery from atomic damage and at temperatures at which this recovery is negligible the optical effects can be removed.

It is hoped to publish a detailed account of this work at a later date.

Acknowledgment

We wish to thank Professor Ditchburn for his encouragement. One of us (E.G.S.P.) wishes to acknowledge the receipt of a maintenance grant from the Department of Scientific and Industrial Research.

References

Berman, R., Klemens, P. G., Simon, F. E., and Fry, T. M., 1950, *Nature, Lond.*, **166,** 864.
Wittels, M., 1953, *Phys. Rev.*, **89,** 657.
Yokota, R., 1952, *J. Phys. Soc. Japan*, **7,** 316.

LETTERS TO THE EDITOR

The Effect of Slip on Sound Propagation

Caro and Martin (1953) have recently investigated the propagation of sound in tubes as a function of the ambient pressure, for pressures down to 5 mm Hg. For air at 20°c in a 5·07 cm diameter steel tube, at a frequency of 1000 c/s, they find a velocity in m sec^{-1} of $(343{\cdot}39 \pm 0{\cdot}04) - (17{\cdot}2 \pm 0{\cdot}45)P^{-1/2} + (3{\cdot}34 \pm 0{\cdot}97)P^{-1}$ where P is in mm Hg. The second term is the ordinary tube correction due to viscous and heat conduction processes. From the term in P^{-1} Caro and Martin calculate the accommodation coefficient to be $0{\cdot}34\,(+0{\cdot}11$ or $-0{\cdot}07)$ on the basis of a formula given by Henry (1931). It is interesting to note that acoustic measurements can be of sufficient precision to allow the estimation of such a coefficient, but actually only the effect of gaseous slip at the tube wall has been considered.

There are, however, two factors which Caro and Martin do not take into account, the term in P^{-1} having at least three components, due to (i) a second-order tube correction, arising because the ratio of tube radius to boundary layer thickness is finite; (ii) the velocity slip at the wall, involving the slip coefficient f' (the accommodation coefficient of Caro and Martin); (iii) the temperature discontinuity at the wall, involving the accommodation coefficient g'. The magnitudes of the above components follow from some work carried out by the writer at Imperial College (see equations (44) and (73) of Weston (1953), the latter equation correcting an error in an algebraic transformation which occurs in Henry's paper). Using the same notation, the theoretical value for the P^{-1} term is

$$c\left[\frac{\gamma'^2}{2r_w^2\omega} + \left(\frac{\pi}{2P\rho}\right)^{\frac{1}{2}}\frac{\mu}{r_w}\cdot\frac{2-f'}{f'} + \frac{(\gamma-1)\gamma^{1/2}k}{r_w\xi}\right]$$

$$= \left[1{\cdot}064 + 0{\cdot}669\frac{2-f'}{f'} + 0{\cdot}453\frac{2-g'}{g'}\right]P^{-1} \text{ for } r_w = 2{\cdot}535 \text{ cm},$$

$$\omega = 2\pi \times 1000 \text{ c/s}.$$

If this is equated with the experimental value, and f' and g' supposed equal (without justification) to enable a value for them to be calculated, it is found that $f' = g' = 0{\cdot}66\,(+0{\cdot}26$ or $-0{\cdot}15)$. This higher figure appears more likely to the author.

The above comments do not affect Caro and Martin's final figure for the velocity of sound in air, since in effect this is determined from an intercept.

Department of Physics,
Imperial College, London.
16th December 1953.

D. E. Weston*.

Caro, D. E., and Martin, L. H., 1953, *Proc. Phys. Soc.* B, **66**, 760.
Henry, P. S. H., 1931, *Proc. Phys. Soc.* **43**, 340.
Weston, D. E., 1953, *Proc. Phys. Soc.* B, **66**, 695.

* Present address: 50 Chiltern Drive, Surbiton Surrey.

The Elastic Constants and Density of Palladium Silver Alloys

Although values of the elastic constants for many alloy systems are to be found in the literature, it is surprising how incomplete are the data. Measurements have usually been made of only one elastic constant and, even where the same observer has investigated more than one, different specimens have often been employed in the work. As a consequence the calculation of any physical property which depends upon the values of more than one elastic constant, such for instance as the Debye characteristic temperature θ_D, is rendered very uncertain.

It has been thought, therefore, that the results of an exploratory investigation of Young's modulus, rigidity and density for a series of palladium–silver alloys, including the pure metals, might be of interest.

The alloys used were kindly lent to us by Dr. B. R. Coles of Imperial College in the form of wires 1·5 mm in diameter. The rigidity of the wires was determined at room temperature, about 18°C, by Maxwell's needle after they had been carefully annealed. Attempts were made to measure Young's modulus by Searle's method, by static flexure and, after drawing down to 0·6 mm diameter, by stretching. None of these methods yielded consistent results to the accuracy considered desirable. Ultimately the wires were helically coiled after reannealing and Young's modulus determined by angular oscillation. In evaluating these results corrections for the finite inclination of the turns to the horizontal were made using the mathematical investigation of Thomson and Tait (1883). These corrections were of the order of 1%. The results obtained are shown in figure 1.

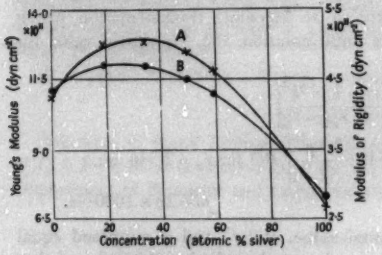


Figure 1. A, Young's modulus and B, modulus of rigidity plotted against concentration for palladium–silver alloys.

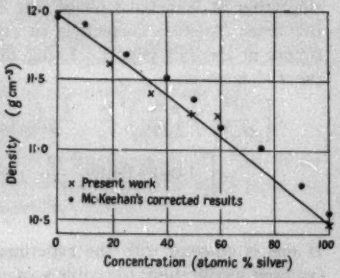


Figure 2. Density plotted against concentration for palladium–silver alloys.

The general shape of the curve for Young's modulus is similar to that obtained by Köster and Rauscher (1948) although we find the maximum to occur at a lower silver concentration than is shown by their curve. The numerical magnitudes are also lower in our results. We have been unable to find any previous results for the rigidity of these alloys.

The densities of the wires were determined using a specific gravity bottle and the results are shown in figure 2 together with the corrected results of McKeehan (1922). McKeehan obtained the densities from x-ray measurements, and in reducing his results used a value of Avogadro's constant of $6{\cdot}0594 \times 10^{23}$; for the corrected results the values have been recalculated using the value

$6{\cdot}0228 \times 10^{23}$ given by Birge (1941). The higher results for the densities from x-ray measurements are to be expected.

From our results the values of θ_D as a function of concentration have been calculated and are shown in figure 3 together with the values obtained from the

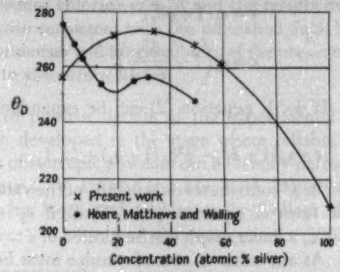


Figure 3. Values of the Debye characteristic temperature θ_D plotted against concentration for palladium–silver alloys.

low temperature investigations of Hoare, Matthews and Walling (1952). This comparison between the curves is offered without comment at the moment since the significance of θ_D calculated from elastic constants is problematical (see Blackman 1941) and the specific heat values may require revision when experiments now in hand at liquid helium temperatures are completed.

The Physics Laboratories,
University of Leeds.
30th October 1953.

F. E. Hoare.
B. Yates.

Birge, R. T., 1941, *Rep. Progr. Phys.*, **8**, 90.
Blackman, M., 1941, *Rep. Progr. Phys.*, **8**, 11.
Hoare, F. E., Matthews, J. C. and Walling, J. C., 1952, *Proc. Roy. Soc.* A, **216**, 502.
Köster, W. and Rauscher, W., 1948, *Zeit. für Metallkunde*, **39**, 111.
McKeehan, L. W., 1922, *Phys. Rev.*, **20**, 424.
Thomson, W. and Tait, P. G., 1883, *Treatise on Natural Philosophy*, §§ 604–608.

A Method of Measuring the Dielectric Constant of Isotropic Powders

Many formulae (see for instance Bruggeman 1935, Van Vessem 1947) have been suggested for the calculation of the dielectric constant ϵ of a dispersion of isotropic particles of dielectric constant ϵ_1 and volume fraction δ_1 immersed in a medium of dielectric constant ϵ_2. Unfortunately none of the formulae can be regarded as better than a first approximation to the actual conditions, and hence any method based upon such formulae is not very suitable for the determination of ϵ_1. The usual alternative method, due to Starke (1897), is to vary ϵ_2 until it exactly equals ϵ_1 : under these conditions the capacity of the measuring condenser is the same whether it is full of the dispersing fluid alone or contains a mixture of the powder and the dispersing fluid. This has been found to be a rather tedious method owing to the need for repeated measurements until $\epsilon_1 = \epsilon_2$. Both methods have been used by various workers, but neither seems to be really suitable for use

as a standard method for the determination of the dielectric constant of isotropic powdered material.

By considering laminae perpendicular and parallel to the field, Wiener (1912) has shown that all mixture formulae should lie between the limiting cases :

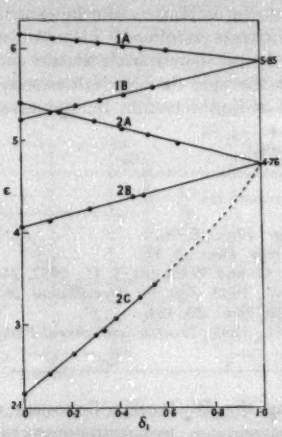
$$\epsilon = \epsilon_2(1+\phi\delta_1) \qquad \ldots\ldots(1)$$

$$\epsilon = \epsilon_2\left(1-\frac{\phi\delta_1}{\phi+1}\right)^{-1} \qquad \ldots\ldots(2)$$

where $\epsilon_1 = \epsilon_2(1+\phi)$. If $\phi<1$, equation (2) can be expanded into the following series:

$$\epsilon = \epsilon_2[1+\phi\delta_1-\phi^2\delta_1(1-\delta_1)+\ldots] \qquad \ldots\ldots(3)$$

When $\phi = 0{\cdot}1$, the term in ϕ^2 contributes only 0·25% of the value of ϵ and the series reduces to the simple form of equation (1). Thus, if ϵ_2 is adjusted to differ from ϵ_1 by less than 10%, a linear graph can be plotted of ϵ to δ_1 and extrapolated to $\delta_1 = 1$ where $\epsilon = \epsilon_1$. As any other mixture formulae must lie between Wiener's extreme formulae, equation (1) offers the basis for an accurate method of determining the dielectric constant of an isotropic powdered material independent of the particle shape.



Variation of dielectric constant ϵ with volume fraction of solid δ_1.
1A, sodium chloride in ethyl acetate ($\phi = -0{\cdot}049$); 1B, sodium chloride in amyl acetate* ($\phi = +0{\cdot}121$); 2A, Pyrex glass spheres in amyl acetate* ($\phi = -0{\cdot}122$); 2B, Pyrex glass spheres in amyl acetate–carbon tetrachloride mixture ($\phi = +0{\cdot}049$); 2C, Pyrex glass spheres in carbon tetrachloride ($\phi = +1{\cdot}128$).

* Different samples of commercial amyl acetate were used.

Preliminary experiments have been performed using Pyrex glass spheres, sodium chloride and potassium chloride immersed in suitable organic liquids or mixtures of organic liquids. Results obtained for sodium chloride and glass spheres are shown in the figure, from which it will be seen that when ϕ is small equation (1) is obeyed. For the glass spheres immersed in carbon tetrachloride

($\phi = 1{\cdot}1$) the graph is not a straight line but a curve, as would be expected from the various mixture formulae. Further, the straight lines intersect at $\delta_1 = 1$, giving values of $\epsilon_1 = 5{\cdot}85$ for sodium chloride and 4·76 for Pyrex glass. Further experiments were performed on potassium chloride, but are not included in the figure because ϵ_1 for potassium chloride is 4·78 and the results overlap those for glass. These values are quite consistent with the published data, but uncertainty in the calibration of the condenser and the possibility of the presence of traces of moisture prohibit any claim to absolute accuracy.

It is intended to publish details of the experimental procedure when the technique has been developed to the stage where reliable measurements of the dielectric constants of isotropic powders can be readily performed. The author is indebted to Professor L. S. Palmer and Mr. J. M. Hough for their guidance and encouragement in his work and to University College, Hull, for a Research Scholarship.

The Physics Department,
University College, Hull.
26th November 1953.

J. A. Reynolds.

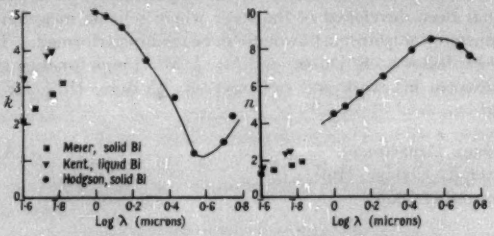
Bruggeman, D. A. G., 1935, *Ann. Phys., Lpz.*, [5], **24**, 636.
Starke, H., 1897, *Wied. Ann.*, **60**, 629.
Van Vessem, J. C., 1947, *Thesis*, University of Utrecht.
Wiener, O., 1912, *Abh. Math. Phys. Kl. Sächs. Akad. Wiss., Lpz.*, **32**, 509.

The Infra-Red Properties of Bismuth

A reflection method of measuring optical constants has been described by Avery (1952) which is suitable for infra-red measurements. A similar method has been used to measure the infra-red properties of bismuth. The method depends on the measurement of the ratio of reflectivities for radiation polarized parallel and perpendicular to the plane of incidence. The values of n, the refractive index, and k, the absorption coefficient, can be calculated from values of this ratio for two angles of incidence θ; n and k are related to the complex refractive index N by $N = n - ik$. Explicit, approximate formulae were used for this calculation, instead of the graphical method used by Avery. The approximation made in the formulae is that $\sin^2\theta$ may be taken as constant. The approximation is better than the experimental error when $\theta \geqslant 70°$. The values $\theta = 74°$ and $78°$ were used for the measurements on bismuth.

Earlier measurements of the infra-red transmission of bismuth films had shown that the transmission increased with increasing wavelength. This phenomenon has been noted by other workers (Schulz 1953). Reflection measurements on bismuth films were possible only for wavelengths less than three microns, because the films were too transparent at greater wavelengths. A piece of bismuth a few millimetres thick was polished on one side to serve as a reflecting surface for infra-red measurements. Molten bismuth was poured on to a cold glass plate and removed when solid. The surface of the bismuth which had been in contact with the glass was smooth except for a few pits. This surface was lightly

polished with metal polish and it retained its polish indefinitely. The thickness of the bismuth was sufficient to eliminate any interference by radiation reflected from the back surface. The values of n and k for this bismuth mirror, for wavelengths between one and ten microns, are shown in the figure. The rise in k at wavelengths less than four microns may be spurious, since the large values of n make the measured values of k less accurate. The probable error in the values for n is about $\pm 0{\cdot}3$, while for k it increases from $\pm 0{\cdot}3$ at $\log \lambda \simeq 0$ to $\pm 0{\cdot}5$ at $\log \lambda > 0{\cdot}4$. Included on the graph are some values of n and k in the visible spectrum, found by Meier (1910) for solid bismuth and by Kent (1919) for liquid bismuth.



The experimental results may be interpreted in the following way. The energy gap between the first filled band and the conduction band of bismuth is about 0·3 ev, corresponding to the absorption minimum at about four microns. The absorption of radiation of wavelength greater than four microns is due to electronic transitions within the conduction band. The results of Kent have often been assumed to show that the optical properties of liquid metals are due to absorption by the conduction electrons. The reason for this assumption is the close agreement between the values of the static conductivity calculated from Kent's values for n and k and the electrically measured values. The value for the number of free electrons per atom calculated from the optical constants of liquid bismuth is 5·1. The infra-red properties of solid bismuth show that the influence of the bound electrons is predominant at wavelengths less than four microns. Theoretical calculations of the zone structure of bismuth (Wilson 1936) indicate that the number of free electrons per atom is 0·0013. It seems unlikely that there should be such a wide difference in the properties of liquid and solid bismuth, that the liquid should have 5·1 free electrons per atom and the solid 0·0013. A theory which attributes the optical properties of liquid metals to transitions of bound electrons must, however, explain the agreement between the optical and electrical values of the static conductivity.

The Observatories,
University of Cambridge.
1st October 1953.

J. N. Hodgson.

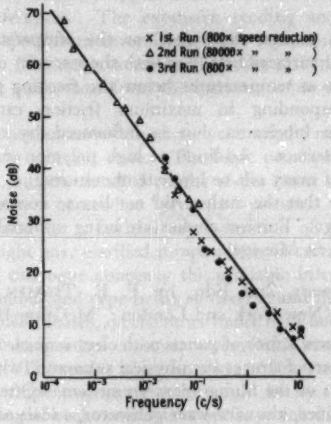
Avery, D. G., 1952, *Proc. Phys. Soc.* B, **65**, 425.
Kent, C. V., 1919, *Phys. Rev.*, **14**, 459.
Meier, W., 1910, *Ann. Phys., Lpz.*, **31**, 1017.
Schulz, L. G., 1953, *J. Opt. Soc. Amer.*, **43**, 406.
Wilson, H. A., 1936, *Theory of Metals* (Cambridge: University Press), p. 86.

Noise in Germanium Filaments at Very Low Frequencies

A recent letter (Rollin and Templeton 1953) reported measurements of current noise in carbon resistors over the frequency-range 7·5 c/s to $2{\cdot}5 \times 10^{-4}$ c/s. A law $\overline{\delta R^2}/R^2 \propto \delta f/f^{1\cdot0}$ was found for the whole range.

The further measurements, on germanium, which were then proposed have now been carried out using the same magnetic recording technique. The specimen used was a centre-tapped, chemically polished filament of 50 ohm cm material, about 12 mm in length and having an end-to-end resistance of about 10 kΩ (cross section 0·6 mm²). All three connections were of the large area soldered type described by Hogarth (1953).

The results differ from those obtained for carbon resistors in that the recording corresponding to the lower frequencies showed a higher playback level than that for the higher frequencies. This is a direct consequence of a noise power per given bandwidth proportional to an inverse power greater than unit of the measuring frequency.



The corrected results are shown in the figure. In this the continuous line corresponds to a law $\overline{\delta R^2}/R^2 \propto \delta f/f^{1\cdot35}$.

Audio-frequency measurements on the same specimen gave a slope of $-1{\cdot}3$.

It should be possible with some minor modifications to use the same technique to measure noise in germanium diodes under reverse bias conditions.

Clarendon Laboratory,
Oxford.
6th January 1954.

B. V. Rollin.
I. M. Templeton*.

Hogarth, C. A., 1953, *Proc. Phys. Soc.* B, **66**, 845.
Rollin, B. V., and Templeton, I. M., 1953, *Proc. Phys. Soc.* B, **66**, 259.

* Now with National Research Council, Ottawa.

REVIEWS OF BOOKS

Reibungsuntersuchen an geschmirten Oberflächen bei tiefen Temperaturen, by M. N. BUTSCHIN. Pp. viii + 64. (Berlin: Akademie-Verlag, 1953.) (German edition of Russian publication 1951.) $1.80.

In recent years much attention has been given to the investigation of materials at low temperatures. The four short papers in this small book emphasize the need for technical achievement to follow up fundamental research. The author is concerned primarily with the possibility of operating mechanisms at low temperatures and discusses apparatus of simple and robust construction suitable for routine testing of bearings and lubricants in the range from +150°C to −120°C.

The first paper considers the influence of physical and chemical surface conditions on metallic sliding from a practical viewpoint and reaches conclusions in general accord with the accepted views in this and other countries, though the author does not appear to be conversant with much of this work.

The subsequent papers show that, as the temperature is reduced, the friction of lubricated surfaces first rises with the viscosity of the fluid, then falls as the solid hardens at temperatures below the freezing point. The 'critical temperature' corresponding to maximum friction can form a basis of comparison between lubricants, but is influenced by the thickness of the layer on the metal surface. Addition of high polymer compounds can extend the effective range of many oils to lower temperatures.

It is unfortunate that the author did not live to continue his work beyond the introductory stage. Further studies are being undertaken by his colleague, R. M. Matwejewski, in Moscow.

G. W. ROWE.

Electronic Measurements, 2nd Edn, by F. E. TERMAN and J. M. PETTIT. Pp. vii + 707. (New York and London: McGraw–Hill, 1952.) 72*s*. 6*d*.

For many years now a rack of panels with electronic circuits behind them has become a standard part of almost any physical apparatus, whether the experiment concerns cosmic rays or the human nervous system. Often enough a physicist can pick up an amplifier, a square-wave generator, a scale of ten counter etc. and build up his apparatus of ready-made units. If he is lucky, there will be as little need to look behind the panel as for the proverbial lady motorist to look under the hood of her car. One must call it fortunate that there is not yet an electronic unit ready-made for every purpose, or that there is not always enough money to buy it, as this encourages a mental laziness which is a serious danger in a highly developed civilization.

The present book by Professors Terman and Pettit of Stanford University, which is the second edition of Terman's work of the same title of 1935, admirably fills the need for an introduction into practical electronics which is elementary, but not a collection of recipes. Almost everything is fully and satisfactorily explained on a level which is accessible to any second year student of physics or engineering, but nothing is detailed to the point where thinking becomes unnecessary. If the student cannot afford to buy more than a dozen books during his whole study this ought to be one of them.

The comprehensiveness of the work is suggested by its (fifteen) chapter headings: Voltage and current, Power, Circuit constants and lumped circuits, Circuit constants in systems involving distributed constants, Measurement of frequency, Waveform, phase and time-interval measurements, Characteristics of triodes, pentodes and similar tubes, Amplifier measurements, Receiver measurements, Antennas, Radio waves, Laboratory oscillators, Generators of special waveforms, Reactance and resistance standards and devices, Attenuators and signal generators.

D. GABOR.

Electron Optics, 2nd Edn, by O. KLEMPERER. Pp. xii+471. (Cambridge: University Press, 1953.) 50*s*.

Like many other subjects electron optics has expanded very considerably in the past fifteen years. The first edition of this book, published in 1938, was one of the earliest dealing with the problems met in focusing electrons. The new edition is nearly five times larger than the first and deals comprehensively with the subject as we know it today.

The book has clearly been written after extensive study of the literature and quotes about 500 references. The extensive reading and reporting in one volume of so many papers from several languages represents the hard toil of a conscientious author. The result is invaluable to the student or research worker as a reference volume quite apart from the intrinsic value of the knowledge set out in its pages.

Electron optics is not an easy subject to write about, mainly because it has still not reached a very coherent state. The basic principles and methods can be stated with clarity, but the detailed results of the many workers in the fields when combined present a somewhat 'scrappy' picture. The author should therefore not be blamed for a faithful representation of this state of affairs; though possibly he might have clarified things a little in certain directions. It is almost impossible to catalogue concisely the available information on electrostatic lenses, their number and type being so diverse and the information being so incomplete. Magnetic lenses, on the other hand, have been investigated more comprehensively and design data to cover a very wide range of lenses can be compactly expressed. The reader who wants to get down to specific lens design is not helped in this respect as much as he might be by the author. This is perhaps a general failing of books on electron optics which tend to deal more with the method of calculating or determining lens properties than with the results so obtained. On the other hand, amongst potential readers a great majority require the results in a form ready for use and have not the facilities for long computations or experimental investigations. The present book does give results in many cases but only somewhat incidentally.

It is evident that the parts of the subject closest to the author's own work are dealt with much more adequately than some others where errors and a number of confusing statements have crept in. This is true, for example, in several places where references to the electron microscope are made. Some recent papers on this subject, although referred to, are not adequately represented. In at least one instance this is unfortunate as the paper has considerable bearing on previous papers fully discussed by the author. Misleading remarks are made about the resistor network, wrongly suggesting it to be very complex and little more accurate than the electrolytic tank.

In general, it can be said that the author is a little uncritical in his writing, and this could lead to misunderstandings in certain places. As an example, the automatic ray tracing method of Sanders, Oatley and Yates is discussed at some length; it is described as a precision method. No indication is given as to the order of precision, although this is of vital importance in the subject. Can the method give adequate precision for the determination of aberration coefficient, or only for accurate focal lengths ?

The book is unique in dealing at length with slit focusing systems and sources and also gives a good description of the increasing number of devices in which electron optics takes an important place.

However many detailed criticisms are made, it must be emphasized that the book is an invaluable source of reference and of general information on the subject. It is written in an easy style with an adequate, but not excessive, amount of mathematics and clearly illustrated with many diagrams. M. E. H.

Practical Thermometry, by J. A. Hall. Pp. 51.
Soft Magnetic Materials Used in Industry, by A. E. De Barr. Pp. 62.
The Magnetic Circuit, by A. E. De Barr. Pp. 62.
Institute of Physics Monographs for Students. (London : Institute of Physics, 1953.) 5*s*. each.

The new Ordinary and Higher National Certificates in Applied Physics, inaugurated by the Institute of Physics in conjunction with the Ministry of Education, are slowly becoming recognized as suitable courses for students whose attendance at Technical Colleges is limited to part-time day and evening classes. Such students will mostly be assistants in the many research and development departments in industrial establishments where physics has become of prime importance, especially since the war.

It is desirable and, indeed, necessary that these students should become knowledgable regarding the applications of physics and that their courses should bear, at least in the final years, a close relation to their professional needs. At the same time it is exceedingly undesirable that these courses should become too specialized, or that the fundamentals should not be stressed at all times in the curriculum. This means that straightforward, not too recondite, presentations of the modern outlook in applied physics, especially regarding materials, instruments and measurement practice, are demanded in their courses.

These first volumes of a series of short monographs, written by specialists for students at Higher National Certificate level, are therefore very welcome. They should also find a place in University Physics Departments, to be read by the degree student especially if he intends to take up a career in industry, since they will guide his thoughts and ideas away from the too severely academic approach.

Practical Thermometry, by Mr. Hall, of the National Physical Laboratory, is a well-written, up-to-date account of the uses, some aspects of the design, and the calibration of electrical resistance thermometers, thermocouples, pyrometers and expansion thermometers. It fulfils admirably the purpose of supplementing the usual textbook accounts of thermometry.

The two monographs by Mr. De Barr, of Messrs. Elliott Bros. (London) Ltd., convey a wealth of information in a limited space, but they are not quite so useful in their approach to the student as Hall's book; doubtless, however, De Barr's

was the more difficult task. In the the volume on *Soft Magnetic Materials*, defined as those where small coercive forces are encountered, an introduction dealing with units, *B–H* and *I–H* curves and associated data, is followed by a short account of the domain theory of ferromagnetism. Apart from a final short chapter on applications, the rest of the text deals with the properties of silicon–iron and nickel–iron alloys. The text has the merits of modernity, originality and authority. But is much of this the sort of information one should convey to students ? Surely in training physicists it is undesirable that they should be expected to assimilate too much purely technical information in graphical and tabulated form, including empirical formulae ? Would it not have been preferable to enlarge considerably the final chapter on applications even at the expense of omitting some of the data on magnetic materials themselves, and even if it meant changing the title of the book ?

The second book, *The Magnetic Circuit*, deals authoritatively with fundamental principles, magnetic powder cores, ferrites and many data on hard magnetic materials and permanent magnets. Again it is felt that, to some extent, too many data have been summarized. For example, it would be better to avoid the complexity of figure 13, which gives no less than ten graphs on one set of axes. For the students' needs, one or two graphs with a fuller account of the reason for the particular choice of axes would have been preferable.

Few authors could have written such a comprehensive account of magnetic materials as Mr. De Barr has done in these little books. Nevertheless, the difficulties of the Higher National Certificate student—often, unfortunately, a man not outstandingly adept at Physics—seem to have been sometimes overlooked. One can visualize the sad lecturer, faced with an audience of young aspirants to learning in physics, having difficulty in conveying the straightforward ideas about the ordinary hysteresis loop, and feeling that details about alloys consisting of an ' eutectoid mixture of cementite and α-iron' are not of much use to him.

These publications of the Institute of Physics are valuable additions to the technical literature available to students and teachers. Is not a word of caution necessary however ? Should not the conveyance of technical information in these booklets, in view of their purpose, be considered as secondary to the demand that they should help the student to understand the subject ? A combination of the industrial expert and the experienced lecturer seems to be necessary in future works.

J. YARWOOD.

Introduction to Geometrical and Physical Optics, by JOSEPH MORGAN. Pp. xi + 450. (New York and London: McGraw–Hill, 1953.) 52*s*.

This is an interesting and well written addition to textbooks available on optics. It covers a wide field from elementary geometrical optics to the quantum theory of spectra and the dualistic aspect of waves and particles. In a book of its size much of the subject must of necessity be treated superficially, but the Foreword explains that the book is intended for students in their sophomore or junior year, i.e. in the second and third years of a four-year course in an American University. A point of interest is that parallel treatments are given wherever possible, the one using and the other avoiding the methods of the calculus. While this may be valuable in the conditions for which the book was

produced, it seems a little unnecessary on this side of the Atlantic, where courses in physics at University level presuppose acquaintance with the elements of the calculus, and include within the first two years of Honours work extensive lectures on differential and integral calculus, including differential and integral equations. The book is consequently more suitable for students reading for a General Honours degree, physics being one of three subjects, and for them it provides an excellent introduction to many ideas which they would find treated in too much detail in those textbooks specially designed for their fellows reading for a Special Physics degree. For these latter students it should prove useful as intermediate reading, serving to whet their appetite for the more exhaustive treatment of advanced textbooks.

The section on geometrical optics uses the satisfactory convention of signs of coordinate geometry, and is clearly and simply written. Refraction by a prism is thoroughly treated, although no mention is made of the curvature of spectral lines. Lens aberrations are clearly explained, stops are carefully treated, and a chapter on photometry includes a brief discussion on the brilliance of optical images. The chapter on optical instruments omits all reference to refractometers other than the prism spectroscope. Full treatment is given to the topics of physical optics, including diffraction by a circular aperture and by a grating, though on page 263 there is a misleading diagram which purports to illustrate the intensity distribution in the primary and secondary diffraction maxima given by a grating. On the opposite page is the correct formula, and reference to this shows immediately that the diagram is in error. The secondary maxima are not all of the same height, but diminish as the distance from the primary maxima increases and, as is well known, the appearance of each primary maximum together with its neighbouring secondary maxima is essentially similar to the Fraunhofer diffraction pattern from a single aperture of width equal to the whole aperture of the grating. As the student is asked to plot the equation, he may find out the mistake, and it is to be hoped that he will have sufficient confidence in himself to disbelieve the printed page.

The treatment of polarized light is good, except for one omission which, while it might appear trifling, is important to the student not already aware of it. Section 16.1 makes reference to section 11.4, but in neither place is it explained that the two arbitrary, orthogonal, linear components into which unpolarized light may be decomposed are incoherent. This is a vital point, and failure to appreciate it will cause difficulty to the student reading section 16.10, in which the composition of two coherent components is correctly shown to give rise to elliptic vibration. The chapter on diffraction gives some account of X-ray spectra, and understandably includes a description of some of the author's researches on the ice-like structure of water. The chapter on spectra gives a good introduction to the subject, though suprisingly no account is given of black-body radiation.

The reviewer has noticed no misprints, but there is an unsatisfactory reference on page 129 where, in a discussion on astigmatism of the eye resulting from a cylindrical defect in the cornea, the reader is referred to section 7.10. This starts on page 105, the greater part dealing with astigmatism produced by oblique passage through a spherical surface, and it is not until the concluding paragraph that astigmatism resulting from cylindrical defect is mentioned. It would have been more helpful had the reader been referred directly to page 108.

There are clear diagrams, and numerous well chosen problems at the end of each chapter. The index is good, and the standard of production is in conformity with that associated with the known excellence of the publisher. Because of the unfavourable dollar exchange the book is expensive, but a student who spends the necessary 52*s*. on its purchase is unlikely to regret having done so.

A. M. TAYLOR.

Progress in Biophysics and Biophysical Chemistry, Vol. 3, edited by J. A. V. BUTLER and J. T. RANDALL. Pp. viii + 386 (London: Pergamon Press, 1953.) 63*s*.

Within recent years there has been an astonishing increase in the number of series of the 'Annual Review', 'Advances', and 'Progress' type. Almost every branch of biological, medical, and physical science is now covered by such periodicals, and indeed there is a danger that in some subjects the number of reviews may soon exceed the number of original papers. It is not surprising therefore that there has been some concern of late as to whether so many books of this type are really necessary. Do they serve to keep scientists informed of progress in other fields, or are their contents written in such a way that they can only be understood by specialists? All too often the latter is the case, and many of these volumes clutter up library shelves without being used to anything like an extent commensurate with their cost.

One sympathizes with the editors of this series in that the selection of material for so ill-defined a subject as biophysics must be very hard indeed. Their view of the subject appears to be that it is largely concerned with the application of physical and physico-chemical techniques to substances of biological origin. This idea can be carried too far. Surely the infra-red spectroscopy of purines or the x-ray crystallography of amino-acids should be regarded as branches of physical science which call for no special knowledge of biology. To take an extreme example, is everyone interested in the properties of water, a substance of the most profound biological importance, to be regarded as a biophysicist?

The professed aim of the editors is to collect articles which may be read with profit by non-experts. Unfortunately this aim has not always been realized, and there must be relatively few people who are even capable of understanding, at least without further study, more than a small fraction of the articles in this series. The present volume contains for example a long article on the theory of the ionic double layer in colloidal systems. There are over 130 mathematical equations, but only five diagrams. The reader is referred to texts of statistical mechanics for the derivation of the very first equation, and formulae are quoted with only a very sketchy indication of their derivation or meaning. One wonders how much the statement that a certain equation of motion can be written in the form

$$\eta \operatorname{curl}^2 \nu + (\nu \,.\, \operatorname{grad} \nu) + \operatorname{grad} p + \rho \operatorname{grad} \psi = 0$$

conveys to most physicists, let alone biologists. (Incidentally, the meaning of the second term, which contains the gradient of a vector, is obscure.) This does not mean that the article is valueless; in the proper place, such as a specialist journal of physical chemistry, it might be excellent, but it cannot by any stretch of imagination be regarded as suitable for the non-specialist. Two other articles, one dealing with the determination of the form and dimensions of particles in solution, the other with transport processes and electrical phenomena in ionic membranes, also tend to be rather mathematical. A certain amount of

mathematics is unavoidable, but it must never be allowed to replace a careful consideration of physical mechanisms. The proper place for it is in original papers, not general reviews. In addition to the papers mentioned, there are others on auto-radiography, ultra-violet microspectrometry, infra-red spectroscopy, viruses and the action of radiations. They are all useful in their way, but make up a very heterogeneous assembly.

This series must inevitably be compared with a similar American production, *Advances in Biological and Medical Physics*, also in its third volume. Already some of the titles are beginning to overlap, and it is doubtful whether there is room for two such series. R. BARER.

Electron Diffraction, by Z. G. PINSKER. (Translated by J. A. Spink and E. Feigl.) Pp. xiv + 443. (London: Butterworths Scientific Publications, 1953.) 63*s*.

This translation of a textbook on electron diffraction by a Russian scientist will be read with a good deal of interest by everyone concerned with this subject. Though necessarily not completely up to date—being published in Russian in 1948—it is the first comprehensive textbook printed since the war.

In its presentation the book follows orthodox lines. It deals, at reasonable length, with the theoretical and experimental aspects of the diffraction of high speed electrons from solids, and gives a detailed account of the experimental techniques used by Russian physicists. Chapters are also devoted to the results of investigations using slow electrons, to diffraction by gases, and to experiments on the polarization of electrons. In fact the book fulfils the requirements of a good textbook on electron diffraction and can be recommended to specialists and non-specialists alike.

Such criticisms of the material and the presentation as can be made could probably also be levelled at any similar textbook which dealt with the experimental and the theoretical side of electron diffraction in a like compass. The presentation of the theory, while in general adequate for most experimental physicists, falls far below that given in the more recent book by von Laue. In particular, dynamical theory is treated in the form originally due to Bethe, and the later work by Lamla in 1938 seems to have been overlooked. In the account of Kikuchi bands and lines, there is no mention of the experiments of Boersch in 1937 on high angle Kikuchi bands. Surprising omissions also occur in the chapter on polish on metals, so much so that the translators have apparently felt obliged to add an important reference which should not have been missed.

To physicists working on electron diffraction the most interesting aspect of this book is the detailed account of work carried out in Russia. Perhaps the most striking instance is the investigation of the relative intensities of the diffraction rings from polycrystalline films of metals. It is found, in contrast to the results of other investigators, that the Mott formula holds for fairly thick films of a large variety of metals, for electrons of energy 15 to 80 kev—to quote one example for films of silver up to 600 Å.

It is concluded, and indeed emphasized, that only kinematical scattering (i.e. single scattering as opposed to multiple scattering) takes place in these films. This conclusion is important as it forms the justification for the application of Fourier analysis to the results of intensity measurements on such films; structure

analysis using such methods constitutes one of the recent developments in electron diffraction—a development which was initiated by Russian physicists.

It is, in this connection, not inappropriate to point out that where the kinematic theory holds, one should expect the thickness of the scattering film to be at least smaller than the elastic mean free path, which can be determined from the Mott formula (together with the number of atoms per cubic centimetre. For silver, and 30 kev electrons, the mean free path is of the order of 35 Å, i.e. about one seventeenth of the value for which it is claimed that kinematic scattering holds. It would appear that more theoretical, if not experimental, work is needed on this aspect of the scattering of electrons in solids.

There is, as is only natural, a tendency throughout the book to stress the importance of Russian contributions to electron diffraction and to physics generally. The well known Bragg law is referred to throughout as the Wulff–Bragg law—a point which the reviewer leaves to others more competent to comment. One would, however, have expected more prominence to have been given to the pioneer work of G. P. Thomson ; most experiments on electron diffraction are carried out by methods which derive directly from Thomson's original work.

The prefaces to the book contain material of a different type, and are interesting as an indication of the general atmosphere existing in scientific circles in Russia. It will come as a surprise to most physicists to learn that there is apparently ' a crisis in bourgeois science ' which has adversely affected *inter alia* the development of certain aspects of electron diffraction.

A special word of praise is due to the translators who have not only produced a very readable translation, but have also added references and the name and the subject index. The production and general presentation of the book leave very little to be desired.

M. BLACKMAN.

Actes du Colloque International des Vibrations non Lineaires, Ile de Porquerolles 1951. Pp. ii+296. (Paris : Publications Scientifiques et Techniques du Ministère de l'Air (No. 281), 1953.) 1,800 fr.

Les modifications de structure du cristal métallique et leur influence sur la cinétique du durcissement structural des solutions solides d'aluminium, by AUREL BERGHEZAN. Pp. iii+95. (Paris : Publications Scientifiques et Techniques du Ministère de l'Air (No. 283), 1953.) 1,200 fr.

Étude théorique et expérimentale de la propagation des intumescences dans les canaux découverts, by JEAN NOUGARO. Pp. ii+155. (Paris : Publications Scientifiques et Techniques du Ministère de l'Air (No. 284), 1953.) 1,600 fr.

CORRIGENDUM

A Note on the Formula for the Mobility of Electrons with Mean Free Path varying with Velocity, by P. M. DAVIDSON (*Proc. Phys. Soc.* B, 1954, **67**, 159).

Equation (1) should read

$$\overline{W} = \tfrac{8}{3}(eE/m)\lambda\overline{c^{-1}}.$$

CONTENTS OF SECTION A

Mr. L. S. Kothari. Reisz Potential and the Elimination of Divergences from Quantum Electrodynamcis : II 201

Dr. J. S. van Wieringen. Justification of the Use of Perturbation Theory in Metallic Conductivity 206

Prof. H. Jones and Mr. B. Schiff. Nuclear Magnetic Resonance in Metallic Lithium and Sodium 217

Mr. R. H. Tredgold. On the Exchange Interaction in the Collective Electron Approximation 221

Dr. W. B. Bonnor. Static Magnetic Fields in General Relativity 225

Prof. H. N. V. Temperley. The Mayer Theory of Condensation Tested Against a Simple Model of the Imperfect Gas 233

Dr. M. E. Rose. Spherical Tensors in Physics 239

Dr. G. L. Squires. The Scattering of Slow Neutrons by Ferromagnetic Crystals 248

Dr. J. R. Prescott. The Decay of ^{203}Pb and the Energy Levels of ^{203}Tl . . 254

The late Dr. W. D. Brodie. Lens Spectrometer Study of the Disintegration of $MsTh_2$ 265

Dr. G. Abraham. Absolute Magnitudes of (d, p) Scattering Cross Sections . 273

Research Notes :

Dr. E. H. S. Burhop. The De-Excitation of Helium Metastable Atoms in Helium 276

Dr. D. E. Alburger and Dr. M. A. Grace. The Disintegration of Cobalt 57 280

Dr. P. F. D. Shaw and Dr. J. R. Prescott. A Search for ^{205}Pb. . . 283

Mr. H. S. Murdoch and Mr. A. J. Webb. Isomerism in ^{46}Ti . . . 286

Dr. N. H. March. On Momentum Distribution in Nuclei . . . 288

Dr. R. E. B. Makinson. The Thermal Conductivity of Metals . . . 290

Dr. L. F. H. Bovey and Mr. W. R. S. Garton. The Absorption Spectrum of Lutetium 291

Letters to the Editor :

Dr. J. H. Crawford, Jr., and Dr. D. K. Holmes. A Chemical Approach to the Treatment of Electronic Spin in Semiconductors 294

Dr. E. W. Elcock. Collective Electron Antiferromagnetism . . . 295

Dr. J. K. Basson and Mr. J. Steyn. Absolute Alpha Standardization with Liquid Scintillators 297

Reviews of Books 299

Contents of Section B 304

Figure 2.

Figure 4.

Figure 5.

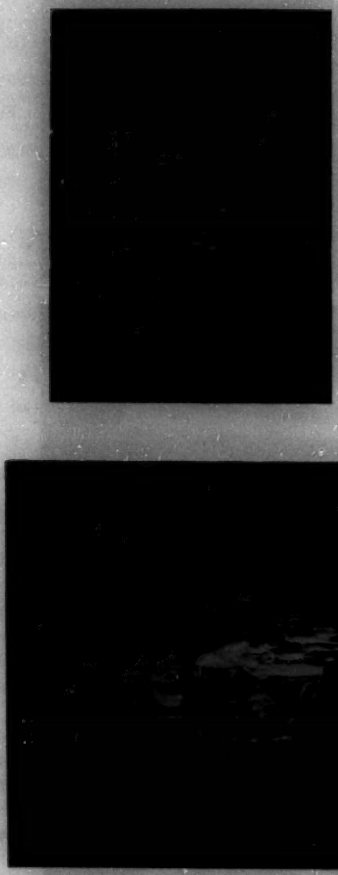


Figure 6. Surface damage caused by sliding stainless steel on stainless steel, (*a*) after a movement of only 25 μ, (*b*) after continued sliding. Normal load 2·8 kg. (Magnification $\times 100$.)

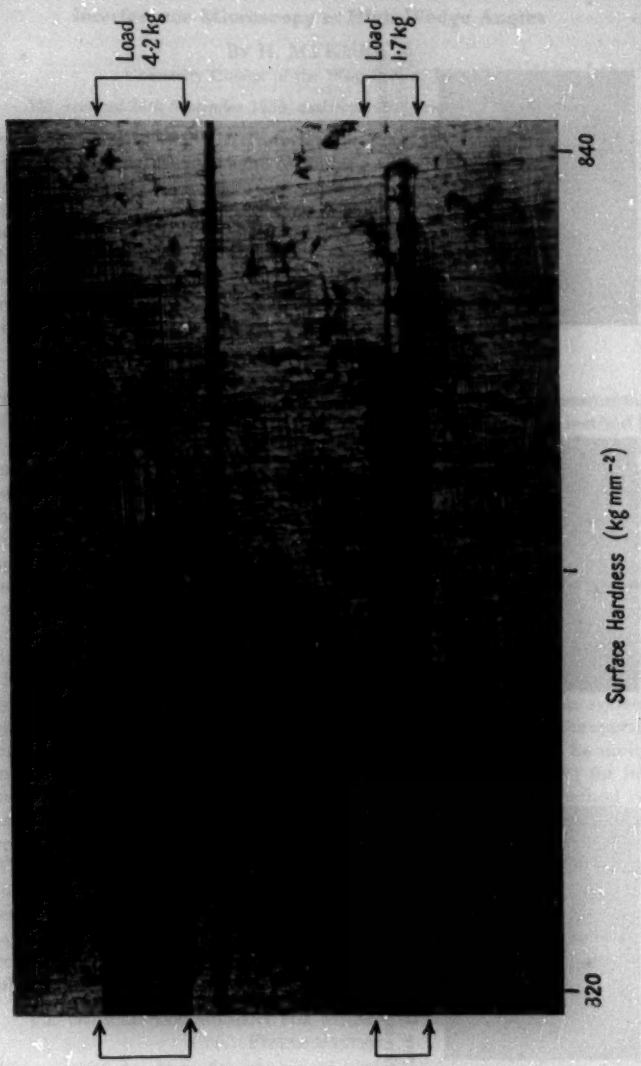


Figure 5. Tracks formed under conditions shown in figure 4.

Figure 1. Oblique illumination, ×750.
Stress = $3{\cdot}2 \times 10^7$ dyn cm^{-2} (tension).

Figure 2. Oblique illumination, ×750.
Stress = $4{\cdot}0 \times 10^7$ dyn cm^{-2} (compression).

Figure 3. Oblique illumination, ×750.
Stress = $6{\cdot}3 \times 10^7$ dyn cm^{-2} (compression).

Figure 4. Phase contrast, ×750.
Stress = $4{\cdot}0 \times 10^7$ dyn cm^{-2} (tension).

Figure 5. Phase contrast, ×750.
Stress = $3{\cdot}5 \times 10^7$ dyn cm^{-2} (compression).

Interference Microscopy at High Wedge Angles

By H. MYKURA
University College of the West Indies, Jamaica

MS. received 28th September 1953, and in revised form 8th December 1953

Abstract. Methods available for interference microscopy of opaque objects at high wedge angles are discussed and formulae for interpreting interferograms are derived. Fringes can be resolved with wedge angles up to 30° both with the Zeiss–Linnik interference microscope and with an ordinary bench microscope using an oil immersion lens and parallel illumination (Fizeau system). The accuracy with which these angles can be measured is poor, about 10% for the highest angle resolvable, but the accuracy increases proportionately with decrease of angle. Some interferograms illustrating the capabilities and drawbacks of the methods are given.

§1. Introduction

Interferometry has been used for surface topography measurement since Fizeau's experiments nearly one hundred years ago. The method has been developed by many workers since then and combined with microscopy to enable examinations to be carried out at all magnifications available to light microscopes. The interference method is most accurate at low wedge angles and becomes progressively less sensitive as the wedge angle increases. However, interference microscopy has many uses for the examination of very small scale (though often large angle) undulations on macroscopically flat surfaces. Examples of such applications are the work of Dew (1952) on diffraction grating replicas and of Hess (1952) on grain boundary grooves. This paper therefore attempts to evaluate the limits of the interferometric method at high wedge angles.

Special interference microscope systems have been produced by Linnik (1933) and Dyson (1953) on the Michelson interferometer principle, where the reference surface is fixed in the instrument. A virtual image of this surface is superimposed on to the specimen so that the optical flat in contact with the specimen under examination, as used in the Fizeau system, is unnecessary. The multiple beam method (Tolansky 1948) is applicable only at low wedge angles; for interferometry at high wedge angles (i.e. greater than 5°) the two-beam method must be used. Tolansky (1952) has also further developed the 'light slit' method, but the maximum accuracy and limitations of this method have yet to be determined.

The Linnik and Dyson systems are particularly useful at high wedge angles, but neither is generally available at present, and no survey of possible systematic errors of these instruments, or of the Fizeau method, when used at high wedge angles, has yet been published. But much can be done with an ordinary bench microscope, using the two-beam Fizeau method, provided certain conditions are observed.

§2. Optical Arrangement for Micro-Interferometry by the Fizeau System

A conventional system for opaque specimens is shown in figure 1. The illumination is arranged so that monochromatic light emerges from the objective as a parallel beam to fall at normal incidence on to the reference flat and specimen.

Monochromatic light is provided by the mercury lamp A and filter C. The image of a small aperture B is projected by lens D and half-silvered mirror E on to the back focal plane F of the objective G, to produce a parallel beam. This falls on to the specimen J through the reference flat H and produces interference fringes which are localized in or near the space between H and J. The object is observed through objective G and eye-piece K acting as an ordinary microscope; and the fringes, which are 'contours' of the surface J with respect to H, are seen superimposed on the object. Feussner (1927) has done much work on the localization of the fringes with respect to the specimen and reference surface and on possible errors of the method, but only at low wedge angles.

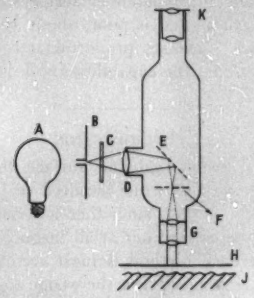


Figure 1. Optical arrangement for micro-interferometry with parallel illumination. See text for detailed description.

If the wedge angle at any point on the object is θ, light on reflection from that point will be inclined at an angle 2θ to the axis of the system. For a ray to re-enter the microscope and form interference fringes the numerical aperture (N.A.) of the objective must therefore be greater than $\sin 2\theta$. If the illuminating beam is not strictly parallel but has a semi-angle of divergence of ϕ this condition becomes N.A. $> \sin 2(\theta - \phi)$. The maximum wedge angle resolvable by different objectives, using this system, can be evaluated from the formula N.A. $= \sin 2\theta$. These angles are given for four typical microscope objectives in table 1, which includes similar values for the Linnik system (derived in the next section) and the maximum wedge angle resolvable in practice.

Employing a 16 mm objective, the system shown in figure 1 works quite well for wedge angles up to 8°. The reference surface can conveniently be provided by placing a microscope coverslip on the specimen. The maximum allowable air gap H–J is determined by the degree of collimation of the parallel illuminating beam and can be evaluated by the formula $8t(1-\cos\phi) < \lambda$, where ϕ is the semi-angle of divergence of the illuminating beam and t is the air gap. For a value of $\phi = 2°$, which is easily attainable, t must be less than 0·1 mm.

However, to resolve wedge angles greater than 13° the 4 mm objective is required. For such a lens the introduction of a coverslip introduces some aberration, though this can be overcome by using a biological objective which is corrected for the presence of a coverslip. Much more serious is the collimating condition. Due to the shorter focal length a greater divergence angle must be

used to obtain adequate light intensity. Further, the objective is used in the reverse direction when collimating the incident light, and therefore introduces errors in collimation towards the edge of the field of view. These effects raise the value of ϕ practically obtainable to about 4°. This in turn requires a maximum air gap of 0·02 mm. Due to unevenness in the typical specimen, the separate glass flat can no longer provide the reference surface.

Tolansky and Omar (1952) have developed a technique of providing a reference surface close to the specimen by placing a thin layer of oil on the specimen. The reference surface is then of an unknown shape but the curvature of the surface will be smooth, due to surface tension forces. As high magnifications are used, only very small areas are examined at one time, and it is found that an oil surface of sufficient flatness over the required area can be produced in practice. In evaluating interferograms using this technique, the wavelength in oil, λ/μ, must of course be used. Due to refraction at the oil–air surface, the maximum angle for which fringes are formed is in this case given by $(\text{N.A.})/\mu = \sin 2\theta$.

Another method of producing a reference surface, of restricted utility in some metallurgical applications, is as follows. The metal is cast on to a coverslip and gas bubbles are trapped between the solidifying metal and glass, thus forming the interferometric gap. This method has been used with cadmium and tin (see also figure 7).

§3. The Linnik and Dyson Systems

The difficulty of bringing a reference surface sufficiently close to the surface under examination was solved by Linnik by superimposing a virtual image of a reference plane on to the surface under examination. The arrangement is shown diagrammatically in figure 2. Convergent illumination is used, and this

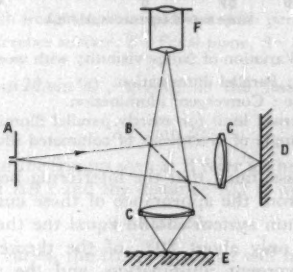


Figure 2. Linnik system (diagrammatic).

A. Light source aperture.
B. Half-silvered reflector.
C. Matched pair of objectives.
D. Reference plane.
E. Specimen.
F. Eyepiece.

means that the 'image' of the reference surface must be within the depth of focus of the objective, coincident with the focal plane. This condition is satisfied during the setting up of the instrument and causes no difficulty. Convergent illumination gives a greater wedge-angle resolving power than parallel illumination for any given N.A., the equation for the maximum theoretically resolvable angle being $\text{N.A.} = \sin\theta$ (see also figure 5 (*a*)).

The firm of Zeiss (Kinder 1937) has produced Linnik type instruments, some with objectives of N.A. = 0·58 having a theoretical angular resolution of 35°. However, these instruments are not readily available.

The Dyson instrument is similar to the Linnik in using convergent illumination. It employs only one objective and has a beam-splitting device between objective and specimen. The instrument can use objectives with N.A. up to 0·85 but is not yet available commercially. A number of other designs have been described (Krug and Lau 1951) but the Zeiss–Linnik instrument may be taken as the best yet produced* for work at high wedge angles.

Convergent illumination can be employed with a simple system similar to that in figure 1, and has a theoretical angular resolution twice that of the parallel illumination system. However, the reference plane must be much closer to the specimen surface, so even with the oil film technique it is often preferable to use parallel illumination. If it is assumed that there is no spurious background, the visibility of the fringes at various wedge angles can be estimated from the fraction of the incident light reflected off the specimen that re-enters the microscope system. Figure 3 gives visibility curves so calculated for parallel and

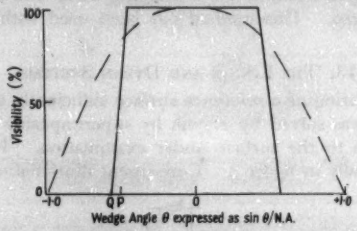


Figure 3. Variation of fringe visibility with wedge angle.
Solid curve : Parallel illumination.
Broken curve : Convergent illumination.
P = Theoretical limit for strictly parallel illumination.
QP = Semi-angle of divergence of collimated illumination beam.

convergent illumination assuming that the interfering beams are equally bright at zero wedge angle. From the appearance of these curves the practical limit for the parallel illumination system should equal the theoretical value, but for convergent illumination only about 80% of the theoretical maximum. The fringe visibility for convergent illumination, and the angular resolution for parallel illumination, can be improved for a given wedge direction by suitably decentring the illumination. This is at the expense of the performance in the opposite wedge direction, but is of use in certain cases.

§4. The Interpretation of Interferograms

When the wedge angle θ is very small the change in air gap between successive fringes is $\lambda/2$ for all systems. The wedge angle θ can be calculated by the formula $\lambda/2l = \tan\theta$, where l is the observed fringe spacing.

* Interference microscopes are now being produced or developed in England by Messrs. Hilger and Watts (Linnik type), C. Baker of Holborn, and Cooke, Troughton and Simms (Dyson type).

(i) *Case of Parallel Illumination* (*Fizeau System*)

Consider now a constant large wedge angle θ with the parallel illumination system, the reference surface being normal to the axis of the microscope. In general, the reference surface, focal plane and specimen surface will be in different positions (figure 4). A ray of light incident through A is reflected at O and crosses the focal plane again at B. In the microscope it is combined with the ray parallel to AO, which is reflected by the reference surface at B′, and the resultant interference fringe appears localized at B. The path difference of the two interfering rays is

$$BB' + AA' + AOB = 2AA' + BC, \qquad \ldots\ldots(1)$$

as $OC = AO$. A similar ray reflected at Q and Q′ will produce a fringe at Q whose path difference is $2AA'$. Therefore the change of path difference for the fringes seen at B and Q is BC. If n fringes are seen betwen B and Q and their spacing is l then $BQ = nl$ and the change in path difference $BC = n\lambda$. But the angle $BCQ = 90°$ and angle $BQC = 2\theta$.

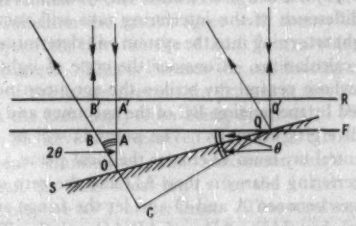


Figure 4. High wedge angle interference with parallel illumination.
R = Reference surface, F = Focal plane, S = Specimen.

Therefore $n\lambda/nl = \sin 2\theta/\sin 90°$, giving as the basic equation for parallel illumination

$$\lambda/l = \sin 2\theta. \qquad \ldots\ldots(2)$$

It must be noted that the fringes seen at B refer to the point O of the specimen. Calling the lateral shift AB s and the distance AO t, we have

$$s = t \tan 2\theta. \qquad \ldots\ldots(3)$$

When the wedge angle varies, the fringe spacing l will be affected by equation (3) as well as (2), the change in lateral shift per order being ds/dn. The use of the $\sin 2\theta$ equation allows for the change in s due to change in t, but a correction should be applied for the change in curvature. So we obtain for the observed fringe spacing

$$l = \frac{\lambda}{\sin 2\theta} + t\frac{d}{dn}(\tan 2\theta). \qquad \ldots\ldots(4)$$

In general the second term of this equation is very difficult to evaluate, but it can cause serious distortion of the fringe pattern when θ is changing rapidly. In practice, errors due to varying lateral shift can be minimized by making t small (i.e. working near points, like Q in figure 4, where the surface is accurately in focus).

(ii) *Case of Convergent Illumination (Linnik System)*

In the Linnik and Dyson instruments the virtual position of the reference plane is made to coincide with the focal plane, so only this case will be treated. Consider a specimen of constant wedge angle θ in an interferometer with objectives of aperture ϕ (figure 5 (*a*)). Of the incident cone of light, only that

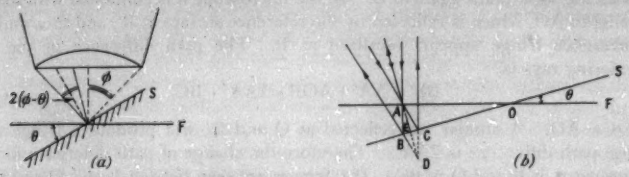


Figure 5. High wedge angle interference with convergent illumination (Linnik system).

portion of angle $2(\phi-\theta)$ marked will re-enter the system. As a convergent beam is used, the path difference of the interfering rays will vary with angle. The central ray of the light returning into the system will determine the fringe position, and is used in the calculations. Consider the cone of light coming to a focus at A (figure 5 (*b*)), whose central ray strikes the specimen normally at E. The cone will be reflected by the portion BC of the specimen and re-enter the optical system apparently diverging from D. The point D will be seen (slightly out of focus) where the central ray from D crosses the focal plane, i.e. at A. The path difference of the interfering beams is then AD and the zero order fringe is at O. Let there be n fringes between A and O and let the fringe spacing be l. Then $AO = nl$ and $AD = n\lambda$, but $2AE = AD$ and $AE/AO = \sin\theta$. Therefore

$$\lambda/2l = \sin\theta \qquad \ldots\ldots(5)$$

for interferometric systems using convergent illumination.

As light from the cone passing through A is reflected from a region BC of the specimen, the fringes give the surface topography accurately only over the depth of focus of the objective. Outside this range the fringes give a mean wedge angle of the portion BC of the specimen, if the wedge angle over this portion is sufficiently constant. This is illustrated in figure 6 (Plate).

A lateral displacement s of the fringes occurs, as in the case of parallel illumination. From figure 5 (*b*)

$$2s = n\lambda\sin\theta. \qquad \ldots\ldots(6)$$

This will also affect the observed fringe spacing when the wedge angle varies. As in equation (4), the observed fringe spacing l for a varying wedge angle becomes

$$l = \frac{\lambda}{2\sin\theta} + \frac{n\lambda}{2}\frac{d}{dn}(\sin\theta). \qquad \ldots\ldots(7)$$

With the Linnik instrument the order of interference n is in practice near zero, so that the second term is usually negligible.

§5. Experimental Tests with Micro-Interferometers

Small steel balls are useful test objects for micro-interferometers. Such balls, of diameter 1·0 to 0·4 mm, are readily obtainable from ball-bearing manufacturers, and their surface condition is adequate for work at highest magnifications.

As test objects they are mounted under a coverslip, which provides the reference surface. The 'Newton's rings' interference fringes are examined, and the known shape of the ball is compared with that deduced from the fringe pattern.

Results of the maximum wedge angles resolvable, as determined by the number of rings visible, are given in table 1. The theoretical values are given for comparison.

Table 1. Maximum Wedge Angle resolved in Ball Bearing Test

Objective	16 mm, N.A. 0·28				8 mm, N.A. 0·45		4 mm, N.A. 0·85		2 mm, (oil) N.A. 1·3	
Test ball diam. (mm)	1·0	0·79	0·39	Theory	0·39	Theory	0·39	Theory	0·39	Theory
Max. angle parallel illumination	7° 50	7° 40	7° 45	8° 00	12° 45	13°	20° 30	29°	22°	45°
Max. angle convergent illumination	11° 40	11° 00	11° 20	15° 40	18°	26° 30	—	58°	—	—

Both the parallel and the convergent illumination results were obtained with a single lens system similar to that shown in figure 1 to test the theory. In the case of parallel illumination the results agree with theory very well for the two lower powered objectives, but for the 4 and 2 mm objective the theoretical resolution is not achieved. This is because the collimating condition (see above) and depth of focus of the lenses prevent the fringes at steep angles (and therefore higher orders with the ball used) from being observed. With other specimens wedge angles up to 30° have been resolved using the 1·3 N.A. oil immersion lens.

With convergent illumination the maximum angle resolved is always appreciably less than the theoretical limit: for the 16 mm objective about 75% and for the 8 mm objective 65% of the limit. This result is to be expected from the contrast–wedge angle curves (see figure 3 above). In the case of the 8 mm the additional reduction is due to the depth of focus of the lens. For the 4 mm lens good fringes can only be expected from about the six lowest orders, therefore the ball bearing test is unsuitable for the high power objectives.

The ball bearing test has also been used on the Zeiss–Linnik instrument. With this no difficulty arises from the order of interference, and fringes are observed at angles up to 30°, the theoretical limit being 35°. Figure 6 is a test interferogram of a 1 mm steel ball taken with the Zeiss–Linnik instrument. The wedge angle in this case varies from 21° to 26°. The fringes are sharp only over the ten lowest orders—at higher orders the fringes no longer contour the surface accurately, as described in § 4(ii) above.

The magnitude of the errors that may arise in the interpretation of interferograms can be judged from table 2. This gives the analysis of a 'Newton's rings' picture of a steel ball 0·79 mm diameter taken through a 4 mm objective, using parallel illumination. The wedge angle expected at various points is calculated from the geometry of the system, first directly, secondly using equation (3), to correct for the lateral shift of the fringes. The wedge angle derived from the fringe spacings is given in line 3 using the formula $\tan\theta=\lambda/2l$ (which is generally used) and in line 4 using equation (2) derived above.

The accuracy of the wedge angles measured interferometrically is to $\pm 2\%$ at 5°, decreasing to $\pm 5\%$ at 14°. The agreement between lines 2 and 4 of table 2, while not perfect, is within the experimental error of the measurements. If the simple formulae are used (comparison of lines 1 and 3, table 2), the error,

while negligible at 5°, amounts to 2° (i.e. three times the experimental error of the interferometric measurement) at 14°.

In the case shown in table 2, the lateral shift of the fringes produces a greater error than the difference between the $\tan\theta$ and $\sin 2\theta$ formulae. If a smaller (0·38 mm diameter) ball is used, the lateral shift causes the smaller part of the total error. The error caused by change in fringe spacing given in equation (4) is negligible in both these cases, but can be very important at a point where the curvature changes rapidly.

Table 2. A Test of the Formulae derived for the Interpretation of High Wedge Angle Interferograms

Order of fringe	6	12	24	36	48
Wedge angle :					
1. Calculated from geometry of ball	5° 05	7° 20	10° 30	13° 00	15° 05
2. As 1 with correction for lateral shift	5° 00	7° 15	10° 10	12° 25	14° 05
3. From fringe spacing by $\tan\theta$ formula	5° 10	6° 55	9° 40	11° 35	13° 05
4. From fringe spacing using equation (2)	5° 10	7° 00	10° 00	12° 05	13° 50

An example of the distortions in fringe pattern due to such a change in curvature is given in figure 7 (Plate). These are interferograms of a grain boundary groove in cadmium, the specimen having been produced by the casting technique mentioned in § 2 above, and photographed through an oil immersion lens using parallel illumination. The wedge angle on either side of the groove is about 23°, so that there is a sudden change in angle of 45° at the base of the groove. In figure 7 (*a*) the focal plane is about coincident with the reference surface, in figure 7 (*b*) about six wavelengths below. The position of focus for best fringe visibility (Feussner surface) does not, in general, coincide with the specimen surface. Serious distortions due to lateral shift, as shown by the difference between figures 7 (*a*) and 7 (*b*), are liable to occur unless care is taken to guard against them.

§ 6. Accuracy of Angle Measurement Attainable in Practice

The accuracy attainable in micro-interferometric measurement of wedge angles, and surface topography generally, will greatly depend on the specimen and apparatus used. It will have all the limitations of the microscope used, plus the errors of measurement of the interference fringes.

Let us consider an ideal specimen which is macroscopically flat within the depth of focus of the microscope objective. With an objective of N.A. 0·58 (as in the Zeiss–Linnik system) and a wedge angle of 30°, ten fringes would be within the depth of focus of the lens. The resolving power of the instrument in this case is nearly equal to the fringe spacing, and though the lateral displacement of the fringes can be corrected for, the fringe system cannot be associated with the object surface to a distance less than the resolving power. The best value for the spacing of ten fringes would then be plus or minus one fringe spacing, or 10%. At smaller angles, where the length of wedge in focus is greater, the

accuracy attainable is correspondingly greater; thus at 10° the maximum accuracy would be 1 in 30 for a wedge 30 wavelengths long. For a wedge of shorter length the accuracy is correspondingly less.

When fringes are widely spaced the difficulty of determining the centre of the fringe arises. In practice it is possible to determine the fringe centre to one-tenth of the fringe spacing, so that for angles greater than 3° the resolving power of the objective, and not the indeterminacy of the fringe centre, limits the accuracy attainable. In general, for any wedge of constant angle θ and observed depth n fringes, the fractional error in determining the depth of the wedge is approximately $0{\cdot}1/n$ and the corresponding error for determining its width (using the resolving power for line objects $\lambda/2\,\text{N.A.}$) is

$$\frac{\text{Lateral resolving power}}{\text{width}} = \frac{\lambda}{2\,\text{N.A.}}\,\frac{2\sin\theta}{n\lambda} = \frac{\sin\theta}{n\,.\,\text{N.A.}}$$

This latter expression predominates at large wedge angles and, therefore, determines the accuracy with which such angles can be measured.

§ 7. Discussion

The comparison of the Linnik and Fizeau systems of interference microscopy shows that both can be used at wedge angles up to 30°. The accuracy with which such wedge angles can be estimated is poor due to the limited resolving power and depth of focus of the microscope. The Zeiss–Linnik instrument is very convenient in use as no separate reference surface need be placed over the specimen. However, due to the high cost of specially built interference microscopes, the Fizeau system as here described may often be more practicable, as an ordinary microscope is used with but slight alteration of the illuminating system.

Acknowledgments

The author wishes to thank the Director of the National Physical Laboratory for the opportunity of using the Zeiss–Linnik instrument at the laboratory and for discussions with members of the staff, also Dr. F. H. Bowen for use of the laboratory facilities of the University College of the West Indies.

References

Dew, G. D., 1952, *J. Sci. Instrum.*, **29**, 277.
Dyson, J., 1953, *Proc. Roy. Soc.* A, **216**, 493.
Feussner, W., 1927, *Gehrcke's Handbuch d. Phys. Optik*, Vol. I.
Hess, J. B., 1952, *Metal Interfaces* (American Soc. for Metals), p. 146.
Kinder, W., 1937, *Zeiss Nachr.*, No. 3, 91.
Krug, W., and Lau, E., 1951, *Die Technik*, **6**, 122.
Linnik, W., 1933, *C. R. Acad. Sci., U.R.S.S.*, 21.
Tolansky, S., 1948, *Multiple Beam Interferometry* (Oxford: Clarendon Press); 1952, *Nature, Lond.*, **169**, 445.
Tolansky, S., and Omar, M., 1952, *Nature, Lond.*, **170**, 81.

Magnetic Transitions and Structural Changes in Hexagonal Manganese Compounds

By B. T. M. WILLIS and H. P. ROOKSBY

Research Laboratories of The General Electric Company, Limited, Wembley, England

MS. received 2nd December 1953, and in amended form 13th January 1954

Abstract. The lattice parameters of the isomorphous compounds MnAs, MnSb and MnBi have been measured over a temperature range which includes the temperatures at which certain magnetic transitions occur. Characteristic changes develop near the magnetic transition temperatures and are of two kinds: (*a*) discontinuous changes of lattice dimensions (MnAs at 40°C, MnBi at 320°C), which are associated with magnetic transitions of the first order, (*b*) discontinuous changes of the temperature derivatives of the lattice parameters (MnAs at 130°C, MnSb at 320°C), associated with second-order magnetic transitions.

Correlation of the results on MnAs and MnSb with those obtained by Greenwald on MnTe indicates that the magnetic exchange energy for manganese atoms at a distance r apart is a maximum for $r = 2{\cdot}94$ Å.

The magnetic change which takes place in MnBi near 440°C is caused by decomposition of the material and is not a Curie point phenomenon.

§ 1. Introduction

All four manganese compounds, MnAs, MnSb, MnBi and MnTe, which crystallize at room temperature with the hexagonal NiAs type of structure (figure 4), show interesting magnetic changes between room temperature and 600°C. These changes consist of transitions between ferromagnetic, antiferromagnetic and paramagnetic states (see table 1), but there is some doubt concerning the precise nature of the transitions in MnAs and MnBi. For convenience we shall denote antiferromagnetic–ferromagnetic, antiferromagnetic–paramagnetic and ferromagnetic–paramagnetic changes by symbols, namely A–F, A–P and F–P respectively.

Table 1. Magnetic States of Hexagonal Manganese Compounds

	Ferromagnetic	Antiferromagnetic	Paramagnetic	Reference
MnAs	< 45°C	? 45–130°C	>130°C	Guillaud (1951 a)
MnSb	<313°C	⟷	>313°C	Serres (1947)
MnBi	<360°C	?360–445°C	>445°C	Guillaud (1951 a)
MnTe	—	<50°C	>50°C	Serres (1947)

MnSb possesses a ferromagnetic Curie point at 313°C, MnTe an antiferromagnetic Néel point at 50°C. The magnetic changes shown by MnAs and MnBi, which are both ferromagnetic at room temperature, are more complex. In the case of MnAs the ferromagnetism disappears at 45°C on heating, but reappears at a lower temperature, 34°C, on cooling. Guillaud (1951 a) has

suggested that this magnetic change is a first-order F–A transition, and that MnAs does not become paramagnetic until a higher temperature, 130°C, is attained. He believes that the magnetic behaviour of MnBi, with corresponding changes at 360°C (or 340°C on cooling) and 445°C, can be similarly explained.

The present work comprises an investigation of the variation with temperature of the lattice parameters of MnAs, MnSb and MnBi. It is shown that characteristic changes of lattice dimensions but not of structure type occur in the neighbourhood of the magnetic transition temperatures. Further, these changes give some indication of the nature of the associated magnetic transitions.

Guillaud (1951 b) has previously reported discontinuous changes of lattice parameter at 45°C for MnAs and 360°C for MnBi, but there are certain significant differences between our results and his. We have also extended the measurements to include the higher temperatures at which the additional magnetic transitions in MnAs and MnBi occur. For MnTe a lattice dilatation at the antiferromagnetic Néel point has been recently noted by Greenwald (1953). All these results are considered in more detail in the discussion below.

§ 2. Experimental Procedure

Powder photographs were taken between room temperature and 600°C using a 19-cm Unicam high-temperature camera and CuKα radiation. This is not the best choice of radiation for manganese compounds, because of secondary fluorescence effects. But since it was necessary to seal the specimens in evacuated silica capillaries, the employment of the softer iron radiation would have required unduly long exposures.

All compounds were prepared by mixing appropriate quantities of the constituent elements and heating in sealed evacuated silica tubes. The preparations of MnAs and MnSb were straightforward. In the case of MnAs, pure manganese and arsenic in the proportions 40 : 60 atomic per cent respectively were well ground together and heated for 20 hours at 800°C. With the exception of one end the tube was lagged, so that during cooling to room temperature the excess arsenic collected at the unlagged end. MnSb was prepared by heating equimolecular proportions of manganese and antimony for one hour at 800°C. Powder photographs of these preparations revealed only lines for compounds having the NiAs-type structure.

The preparation of MnBi presented certain difficulties. According to Adams, Hubbard and Syeles (1952), the formation of MnBi from manganese and bismuth takes place by means of a peritectic reaction. After numerous experiments the most satisfactory preparation was obtained as follows. Manganese and bismuth were mixed in the ratio 45 : 55 atomic per cent and heated *in vacuo* for one hour at 1000°C. The temperature was then lowered slowly to 440°C, held there for several hours, and finally reduced to room temperature. The powder photograph of the product showed the presence of MnBi (NiAs-type structure) and bismuth in roughly equal proportions. Further enrichment of the MnBi content was achieved by magnetic separation. Little confusion in interpretation was caused by residual bismuth in the specimen as the principal x-ray observations were made above the melting point of bismuth. We must emphasize, however, that the results on MnBi described below refer to the alloy in the presence of excess bismuth.

§ 3. Results

The x-ray data are conveniently summarized in the lattice parameter–temperature curves, figures 1–3.

MnAs

As the temperature rises through 39°c (figure 1) a discontinuous decrease of *a* takes place, leading to a contraction of structure cell volume of 2%. No corresponding change takes place in the *c*-axis direction. The precise magnitudes of these lattice changes are recorded in table 2.

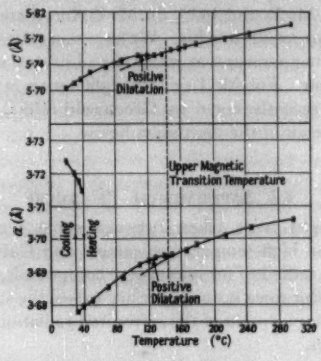


Figure 1. Curves showing variations with temperature of structure cell dimensions of MnAs.

Table 2. Lattice Dimensions of MnAs, MnSb and MnBi

	MnAs			MnSb	MnBi		
		Just below	Just above			Just below	Just above
Temp. (°c)	20	39	39	20	20	320	320
a(Å)	3·724	3·715	3·681	4·128	4·286	4·310	4·380
c(Å)	5·706	5·721	5·721	5·789	6·116	6·182	5·991
$V^{1/3}$ (Å)		4·090	4·064			4·634	4·634

By careful adjustment of the temperature of the two furnaces it was found possible to produce a temperature gradient of a few degrees across a specimen at the mean temperature of 39°c. In this way diffraction lines of both low- and high-temperature 'structures' were recorded on the same film. The two sets of lines could easily be distinguished, as they were inclined at a small angle to each other. All lines were sharp and clearly defined, showing that the lattice change was strictly discontinuous.

The change exhibited small thermal hysteresis. On cooling the material from above 39°c the lines of the low-temperature structure failed to re-appear until a temperature of 32°c was reached.

These results are in good agreement with those of Guillaud (1951 a), who measured the lattice parameters at two points on the heating curve, at 20°c and 45°c. Minor differences between the two sets of results, such as small differences in the absolute values of lattice dimensions and of transition

temperatures, can probably be accounted for on the basis of a small difference of composition.

The second region of interest in figure 1 is in the neighbourhood of 140°c, where both the (a, T) and (c, T) curves show an inflection. The lattice dilatation involved here is characteristic of an A–P change. For other antiferromagnetic compounds with rhombohedral or hexagonal symmetry, e.g. Cr_2O_3 (Smart and Greenwald 1951), α–Fe_2O_3 (Willis and Rooksby 1952) and CrSb (Willis 1953), the anomalous dilatation occurs principally along the c-axis direction. But there appears to be no reason in principle why the dilatation should not take place normal to this axis as in MnAs. Thus the x-ray data support the suggestion of Guillaud (1951 a), based on magnetic evidence, that an A–P change occurs at approximately 130°c.

MnSb

The (a, T) curve (figure 2) shows a pronounced inflection near 320°c, whereas the (c, T) curve is of nearly uniform slope throughout the entire range from room temperature to 600°c. The ferromagnetic Curie point is 314°c.

MnBi

There is a discontinuous change in both a and c at 320°c (figure 3), a increasing by 1·5% and c decreasing by 3% on heating through this temperature. The resulting change of cell volume is less than 0·1% (table 2).

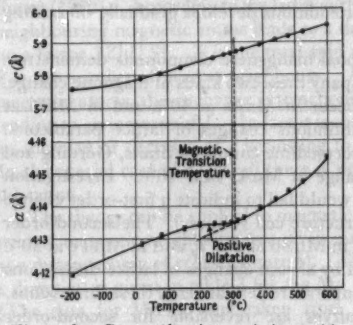


Figure 2. Curves showing variations with temperature of structure cell dimensions of MnSb.

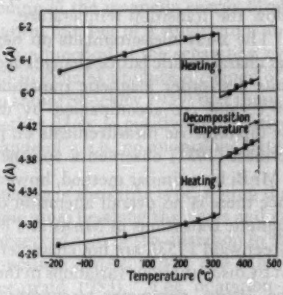


Figure 3. Curves showing variations with temperature of structure cell dimensions of MnBi.

We are unable to explain completely the differences between these results and those of Guillaud (1951 b), who was the first to observe this dilatation. Guillaud made x-ray measurements at room temperature on two samples, one of ferromagnetic MnBi cooled slowly from above the transition, and the other of non-ferromagnetic MnBi rapidly quenched from 400°c. He deduced that c decreases by 5% whilst a remained unchanged on heating through 360°c. There are thus differences both in the nature of the dilatation and the temperature at which it occurs. On the other hand, our results are in good agreement with the observations of Adams, Hubbard and Syeles (1952), who give $c = 5{\cdot}87 \pm 0{\cdot}02$ Å for MnBi rapidly quenched from 320°c.

On heating samples of MnBi beyond 320°C we observed an irreversible change of structure in the region of 435°C. The diffraction lines of the NiAs-type structure were replaced by lines of an unidentified phase, and this phase persisted on subsequent cooling below 435°C. The existence of a magnetic change at 440°C has been known for some time. Fürst and Halla (1938) attributed it to a Curie point change, whereas other authors (Bekier 1914, Siebe 1919) considered that a peritectic reaction takes place at 440°C. More recently Guillaud (1951 a) and Smart (1953) have lent support to the former point of view in suggesting that 440°C represents the antiferromagnetic Néel temperature of MnBi. The present results show clearly that decomposition of MnBi occurs before the Néel temperature, if it exists, is reached, and that there is therefore a fundamental difference between the behaviour of MnAs and MnBi at their upper magnetic transition points (130°C and 440°C respectively). The magnetic change in MnAs is an order–disorder effect, which involves no structural change apart from a small adjustment of lattice parameters; in MnBi the magnetic change is caused by decomposition.

§ 4. General Discussion

Magnetic changes taking place without alteration of composition or phase can be broadly classified into transitions of the first or second order (Smart 1953). The former include F–A transitions and give rise to a discontinuous change of orientation of electron spins. Second-order transitions include F–P and A–P changes; in these an ordering of spin orientations develops gradually on cooling below the transition temperature.

The x-ray measurements on hexagonal manganese compounds demonstrate that characteristic lattice changes accompany these two kinds of magnetic change. Thus first-order magnetic transitions occur in MnAs at 40°C and in MnBi at 320°C, and are accompanied by discontinuous changes of lattice parameters. (By dilatometric measurements on polycrystalline compacts Smits, Gerding and Vermast (1931) observed a volume change in MnAs near 40°C. Investigation of MnBi by a similar method, however, would fail to indicate a first-order change since there is no overall alteration of structure cell volume.) The second-order magnetic transitions in MnAs at 130°C, in MnSb at 314°C, and in MnTe at 50°C (Greenwald 1953), are not characterized by abrupt changes in lattice dimensions but by discontinuous variations in the temperature derivatives of these dimensions. The lattice parameter–temperature curves are reversible for second-order changes; in the case of first-order changes there is thermal hysteresis in the transition region.

Occasionally it is even possible to distinguish between second order F–P and A–P changes by purely x-ray means. F–P transitions in cubic substances (e.g. α-Fe at 770°C) take place without change of lattice symmetry, whereas A–P transitions in materials which are cubic in the paramagnetic region (e.g. NiO, MnO, CoO) lead to a slight deformation to rhombohedral or tetragonal symmetry below the transition point. No change of symmetry occurs in non-cubic substances (e.g. MnSb, MnAs, MnTe) and the above distinction does not arise.

Smart (1953) has considered theoretically the dilatation accompanying first-order magnetic changes. He suggests that the magnetic atoms in the majority of antiferromagnetic materials are arranged in sheets of atoms of the

same spin, so that in an A–F change a dilatation takes place normal to these sheets. Thus in MnBi he considers the sheets to be normal to the *c*-axis, and in MnAs to contain the *c*-axis. Unfortunately, his model for MnBi fails to account for the abrupt change in both *a* and *c*; moreover, for MnAs one would expect a symmetry lower than hexagonal in the antiferromagnetic region, whereas the x-ray results give no evidence of such an effect.

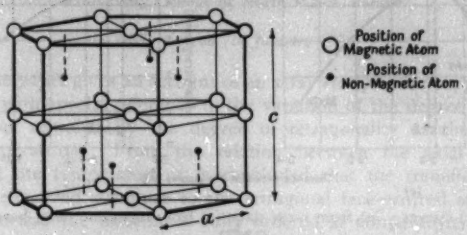


Figure 4. NiAs-type crystal structure : the magnetic atoms occur in plane sheets, normal to the *c*-axis and of spacing $c/2$.

Theoretical treatments of the thermal expansion associated with second-order changes have been given by Néel (1937), Shockley (1939), and Greenwald and Smart (1950). According to the latter authors the exchange interaction J between neighbouring magnetic atoms leads to a deepening of the magnetic energy levels by a factor $2Z|J|S^2$. S is the total spin quantum number of a single atom and Z is the number of nearest neighbours. As J is a function of the separation r of magnetic atoms, the exchange interactions will in general produce a change of r leading to an increase of J. In fact it is easy to show (Shockley 1939) that there is an increase or decrease of r according to whether $\partial J/\partial r$ is positive or negative. In the special case of $\partial J/\partial r = 0$ there is no change of r, i.e. no anomalous lattice dilatation.

The theory can be applied to magnetic materials of the NiAs type, noting that nearest magnetic neighbours to a given magnetic atom lie along the *c* axis and are at a distance away of $c/2$ (see figure 4). The observed sign of the dilatation in the *c* direction is given in table 3 for the Mn compounds MnAs, MnSb and MnTe. (The data for MnTe were taken from Greenwald 1953). The dilatation is zero for MnSb and we can conclude that $\partial J/\partial r = 0$ for $r = \frac{1}{2}c = 2{\cdot}94$Å, where J is the exchange interaction between neighbouring Mn atoms; $\partial J/\partial r$ is positive for $r = 2{\cdot}88$Å and negative for $r = 3{\cdot}36$Å. From these data the curve in figure 5 (*a*), relating J and r, has been constructed. J is given in arbitrary units and the ordinates corresponding to the points for MnAs and MnTe have been estimated roughly, assuming that $\partial J/\partial r$ is proportional to the observed magnitude of the dilatation.

It is of interest to compare figure 5 (*a*) with figure 5 (*b*), which represents the (J, r) curve for Mn as deduced by Guillaud (1943) using an independent method. Guillaud's method consisted of plotting the second-order transition temperature T_c against r. T_c can be taken as a measure of J (in arbitrary units), as T_c is related simply to J by the equation $T_c = Z|J|2k$, where k is Boltzmann's constant. There is serious doubt as to the validity of including the point for MnBi in

figure 5 (*b*). Guillaud assumed that MnBi possessed a second-order transition point at 440°c, whereas the present work indicates that there is instead a complete transformation. Figure 5 (*b*) contains an additional point for MnTe, which was not included by Guillaud. The correspondence between figures 5 (*a*) and (*b*)

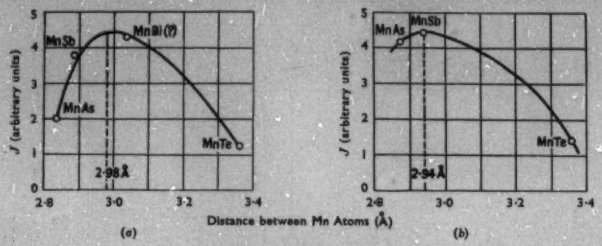


Figure 5. (*J*, *r*) curves : (*a*) from x-ray data, (*b*) from magnetic data, based on Guillaud (1943).

appears to be fairly good. Both curves indicate that *J* passes through a maximum value (in accordance with the theoretical considerations of Bethe, see Bozorth 1951), and the closeness of the two estimates, 2·94Å and 2·98Å, for the position of the maximum is noteworthy. Unfortunately the two curves possess too few points. Additional points could be obtained by taking observations on materials with different *c* parameters, and the preparation of solid solutions between the several compounds may well provide the required range of *c* values.

Table 3. Lattice Dilatation at Second-Order Transition

Compound	$r\ (=c/2)$	Obs. sign of dilatation	Sign of $\partial J/\partial r$
MnAs	2·88Å	+ve (figure 1)	+ve
MnSb	2·94Å	0 (figure 2)	0
MnTe	3·36Å	−ve	−ve

References

Adams, E., Hubbard, W. M., and Syeles, A. M., 1952, *J. Appl. Phys.*, **23**, 1207.
Bekier, E., 1914, *Intern. Z. Metallographie*, **7**, 83.
Bozorth, R. M., 1951, *Ferromagnetism* (New York : Van Nostrand).
Fürst, U., and Halla, F., 1938, *Z. phys. Chem.* B, **40**, 285.
Greenwald, S., 1953, *Acta Cryst., Camb.*, **6**, 396.
Greenwald, S., and Smart, J. S., 1950, *Nature, Lond.*, **166**, 523.
Guillaud, C., 1943, *Thesis*, University of Strasbourg.
Guillaud, C., 1951 a, *J. Phys. Radium*, **12**, 223; 1951 b, *Ibid.*, **12**, 143.
Néel, L., 1937, *Ann. Phys., Paris*, 11, **8**, 237.
Serres, A., 1947, *J. Phys. Radium*, **8**, 146.
Shockley, W., 1939, *Tech. Publ. Bell Tel. Syst.*, **18**, 645.
Siebe, Z., 1919, *Z. anorg. Allgem. Chem.*, **108**, 161.
Smart, J. S., 1953, *Phys. Rev.*, **90**, 55.
Smart, J. S., and Greenwald, S., 1951, *Phys. Rev.*, **82**, 113.
Smits, A., Gerding, H., and Vermast, F., 1931, *Z. phys. Chem.* (Bodenstein Festband), p. 357.
Willis, B. T. M., 1953, *Acta Cryst., Camb.*, **6**, 425.
Willis, B. T. M., and Rooksby, H. P., 1952, *Proc. Phys. Soc.* B, **65**, 950

Experimental Study of the Variation of the Degree of Order with Temperature in a Copper–Palladium Alloy

By D. MADOC JONES and E. A. OWEN
University College of North Wales, Bangor

MS. received 7th January 1954

Abstract. The paper gives an account of an x-ray study of the copper–palladium alloy Cu_3Pd, with special reference to the variation of the degree of order with temperature as revealed by the degree of tetragonality of the structure at different temperatures. From the relation between the axial ratio of the structure and the temperature it is concluded that the transformation from the cubic face-centred structure to the tetragonal face-centred structure takes place at 460°c and the material is fully ordered at temperatures below about 270°c when the axial ratio c/a is 0·9851 and the a-parameter 3·7021 kx at 18°c. On the assumption that the degree of tetragonality gives a measure of the degree of order it is concluded that the relation between the axial ratio and the degree of order is given by $c/a = 1 - 0{\cdot}015W$ when W is the degree of order.

§ 1. Introduction

The copper–palladium alloys have been investigated by Jones and Sykes (1939), particular attention being paid to the alloys Cu_3Pd and CuPd with a view to determining how far the transformations observed in these alloys verified the predictions of order–disorder theory. The investigation was conducted by measuring the electrical resistivity and the specific heat as well as by taking x-ray photographs of the alloys after various heat treatments. It was found that slowly cooled Cu_3Pd is in the ordered state and has a face-centred tetragonal structure whereas at temperatures above about 460°c it is disordered having a face-centred cubic structure. It was concluded also that the atoms do not arrange themselves as in the ordered structure of Cu_3Au. But if the structure were considered in terms of the disordered face-centred cell existing at high temperature then as the cooling proceeded it was found that the axial ratio a/c of the tetragonal structure below about 460°c diminishes from 1·000 for the disordered material to 0·986, with $a = 3{\cdot}707$ Å, for material which has been slowly cooled over four months to room temperature.

In the present paper a study is made of the alloy Cu_3Pd with a view to obtaining information concerning the degree of tetragonality as a function of the temperature, and thereby the variation of the degree of order with temperature.

§ 2. Preparation of Specimens

Three alloy ingots were prepared from spectroscopically pure palladium and copper, each alloy having a composition of 25 atomic per cent palladium and 75 per cent copper, and each weighing approximately 2 g. The ingots were examined for segregation and were found to be satisfactory in this respect; no difficulty was experienced in producing alloy ingots that were homogeneous and of the correct composition within $\pm 0{\cdot}1$ atomic per cent.

Two different types of investigation were carried out. In the first the alloys were exposed to the x-ray beam when maintained at various temperatures in a specially designed high-temperature focusing camera which will be described elsewhere. This investigation enables the critical temperature T_c to be determined at which the transformation from the face-centred cubic structure to the face-centred tetragonal structure, takes place, that is, the transformation from the disordered to the ordered state. It will also show how the degree of tetragonality depends upon the temperature of the alloy. In the second part of the investigation room temperature x-ray photographs were obtained of specimens quenched from various temperatures. This supplies information similar to the first and serves as a check upon it.

§ 3. High-Temperature Investigation

It was possible with the high-temperature camera employed, by taking special precautions which will be referred to later, to record the variation of the lattice parameter of the disordered alloy not only above the transformation temperature of 460° but also below this down to room temperature. The variation of the parameters of the ordered alloy over the range of temperature from 380°c to 460°c could be examined in terms of the original face-centred cubic cell by studying how the main lines from this face-centred cubic lattice split up as the new structure is formed.

(*a*) *Disordered material.* It was arranged that the material was always in the disordered state before starting observations upon it at any temperature. X-ray photographs taken when the specimens were maintained at temperatures above about 460°c showed a face-centred cubic structure in accordance with the findings of previous workers. The ordering process in the alloy takes place relatively slowly so that by rapidly cooling the specimens to fixed values of temperature below 460°c and taking the x-ray photographs as soon as possible afterwards, it was possible to find the size of the cubic lattice at the lower temperatures. Each exposure lasted 30 minutes and the average cooling period between each photograph was about 60 minutes, so that the total time interval during which the specimen was maintained at any given temperature was comparatively short. The results in the order they were obtained are recorded in the table.

Data for Parameter–Temperature Curve for the Cubic Lattice

Temp. (°c)	Lattice parameter (kx units)	Temp. (°c)	Lattice parameter (kx units)
18	$3{\cdot}6854_6$	452	$3{\cdot}7105_6$
474	$3{\cdot}7123_1$	428	$3{\cdot}7089_6$
506	$3{\cdot}7147_6$	21	$3{\cdot}6847_1$
541	$3{\cdot}7170_3$	377	$3{\cdot}7056_9$
586	$3{\cdot}7203_2$	296	$3{\cdot}7003_2$

In the course of this part of the investigation, when cobalt Kα radiation was used to obtain the x-ray photographs, it was observed that there were slight differences in the parameter values calculated from the (331) and (420) reflections. It was evident that in the short interval during which the material was maintained below the transformation temperature, the change in the position of the lines which occurs when the tetragonal structure is formed, had already begun. This is also shown in the table by the values of the lattice parameter at 18°c and 21°c—two observations made with an interval of some hours between them—

the value of the lattice parameter in the latter observation being lower than that in the former. The true parameter–temperature curve of the cubic phase should therefore pass through the first point obtained at 18°C and through all the points above 460°C whilst it should lie slightly above the points at temperatures below 460°C. On plotting the points this will be found to be the case.

(*b*) *Ordered material.* The lattice parameters of the tetragonal structure present greater difficulty of measurement because the degree of tetragonality depends upon the temperature, and it is essential that the specimen should be in a state of equilibrium in order to give the correct lattice parameter value corresponding to that temperature.

Each specimen was given a preliminary annealing of 12 hours at 540°C to bring it to the disordered state. It was then mounted in the camera and maintained at 470°C for some hours in the evacuated camera chamber. The temperature was then reduced in steps of about 20°C and the specimen maintained at each temperature until consecutive photographs at that temperature gave the same parameter values; it was then assumed to have attained the equilibrium state at that particular temperature. These measurements showed that the rate of change of the *a*-parameter, for example, with period of annealing is rapid in the initial stages, gradually falling to zero as the alloy reaches equilibrium. Such curves were obtained for each temperature.

As the transition temperature T_c is approached from the ordered phase, the axial ratio of the face-centred tetragonal structure approaches the value unity. Hence if the axial ratio–temperature curve is plotted and extrapolated to unit value of the axial ratio, this gives the critical temperature T_c. Also as the axial ratio tends to unity the two parameters *a* and *c* approach each other in value and at the critical temperature become identical and equal to the cubic lattice parameter at that temperature. The point of intersection of the three parameter–temperature curves also gives the critical temperature (figure 1).

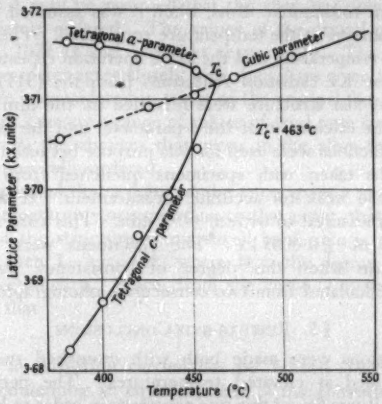


Figure 1. Parameter–temperature curves for the alloy Cu_3Pd in the range 380°C to 550°C

The mean value of the critical temperature found from these curves is 463° ± 3°C, which agrees quite well with previous values.

In order to obtain the specimen in a state of equilibrium, for example at 381°c, it had to be maintained at this temperature in the evacuated camera chamber for 25 days. This put severe strain on the camera; the consistent results obtained, without a breakdown of any kind, prove its suitability for problems of this nature. The focusing camera is more convenient than the rotating-fibre camera for carrying out this type of investigation because accurate values of the tetragonal parameters are directly obtained by a simple graphical method which cannot be applied to films taken on the rotating-fibre camera. The focusing camera has one disadvantage however in that the range of reflection angles is rather restricted and it is not always possible to obtain reflections which lie within this range.

§4. Quenched Specimens

To investigate quenched specimens lattice spacing values were obtained from the measurement of photographs taken at room temperature with a focusing camera which had been so constructed that, during the x-ray exposure, the powder reflector could be oscillated backwards and forwards along the circumference of the camera. This arrangement made it unnecessary to use very fine powder because continuous diffraction lines which could be accurately measured were produced even with coarse grain powder.

Preliminary experiments with quenched specimens and x-ray photographs taken with the rotating-fibre camera confirmed the existence of the transformation from cubic structure at high temperature to tetragonal structure below 460°c. The procedure with quenched specimens was similar to that adopted in the high-temperature investigation; the samples were initially brought to the disordered state, then heated at a given temperature for a certain period and quenched. After taking an x-ray photograph with the specimen it was replaced in the furnace for a further period of annealing of the same duration at the same temperature. This procedure was continued until constant parameter values were obtained from consecutive films, when it was assumed that the specimen had attained equilibrium at the temperature considered. The furnace was then cooled to the next temperature and the same operation repeated.

With manganese Kα radiation reflections from the (311), (113) and (222) planes of the tetragonal structure were recorded on the film. Two lines only are necessary for the calculation of the a-parameter and the axial ratio. Usually (311) and (222) reflections were used for this purpose because the (113) reflection in the photographs taken with specimens quenched from near the critical temperature was too weak for accurate measurement. It was estimated that the arcs could be measured to within $\pm 0{\cdot}1$ mm. This corresponds to a lattice parameter change of $\pm 0{\cdot}0005$ kx. The specimens were assumed to have attained equilibrium when this degree of consistency was shown by the parameter values calculated from two consecutive photographs.

§5. Results and Conclusions

Many observations were made both with quenched specimens and with specimens maintained at elevated temperatures. The periods of annealing were long because it was found this was necessary in order to bring the specimen to a state of equilibrium at a chosen temperature. Since accurate lattice spacings could be obtained with the focusing camera furnished with the powder oscillating mechanism and that the photographs with this camera were taken at room

temperature, the final results were obtained with quenched specimens using manganese Kα radiation. From the reflections recorded on the films the values of the *a*-parameter and the axial ratio of the tetragonal structures were directly deduced from the measurements of the reflections from the planes (311) and (222). The values of the axial ratio are shown plotted against the temperature in figure 2. This curve is similar in shape to the curve of *a*-parameter plotted

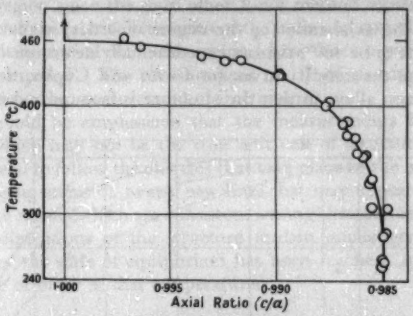


Figure 2. Variation of axial ratio with temperature from the disordered to the ordered state of alloy Cu_3Pd.

against temperature, but the change in the degree of order with temperature is more conveniently studied by means of the axial ratio curve in figure 2 which shows that the critical temperature is 460°c, again in good agreement with previous values. The curve also shows that the axial ratio has a constant value of 0·9851 below 270°c, and it may be assumed that the alloy has now reached a state in which it is completely ordered. The values of the crystal parameters *a* and *c* at 18°c are then 3·7021 kx and 3·6469 kx respectively. The values of the axial ratio and the *a*-parameter agree closely with the values given by Jones and Sykes when the structure is considered in terms of the original face-centred cubic cell. The curve in figure 2 covers a range of values of axial ratio from 1·000 to 0·9851, and therefore covers the changes that occur in the alloy from the disordered state until it becomes fully ordered.

If it may be assumed that a point on the curve gives a measure of the degree of order at the temperature corresponding to that point, that is, that the degree of tetragonality of the structure as measured by $1-c/a$ is proportional to the degree of order, then $1-c/c=kW$, where W is the degree of order and k the constant of proportionality. Since, from the curve, $c/a=0{\cdot}9851$ when $W=1$, then $k=0{\cdot}015$ so that

$$c/a=1-0{\cdot}015W. \qquad \ldots\ldots(1)$$

A theoretical investigation carried out by Wilson (1938) applicable to an alloy like AuCu containing an equal number of the constituent atoms, showed there was a simple relation in this case between the degree of tetragonality and the degree of order, the relation being

$$c/a=1-0{\cdot}06W \qquad \ldots\ldots(2)$$

the coefficient of W having been derived from considerations based on the size of the atoms. The value of the coefficient for CuAu on this basis is therefore about four times that found experimentally for Cu_3Pd. Both these alloys have face-centred cubic structures in the disordered state changing to face-centred tetragonal structures when order sets in at the lower temperatures, but the alloy Cu_3Pd differs from CuAu in that the constituent atoms are present in the ratio of 3 : 1 and not 1 : 1 and would thus appear to present greater difficulty of calculation.

In figure 3 the axial ratio (or the degree of order) is plotted against T/T_c where T_c is taken to be 460°, the last experimentally determined value. Figure 3 also includes the theoretical curves for CuAu and Cu_3Au, the latter differing from the other two alloys in that the structure is face-centred cubic both in the

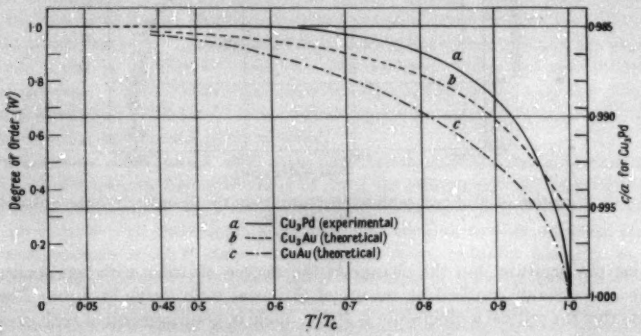


Figure 3. Relation between the degree of order W and the temperature expressed as a fraction of the critical temperature T_c.

ordered and in the disordered states, and its curve shows that when the critical temperature is approached a certain degree of order remains which suddenly disappears when the critical temperature is reached. This discontinuity is not observed to any extent in the curves of the other two alloys, CuAu and Cu_3Pd, the degree of order decreasing slowly at first but afterwards more rapidly and continuously as the critical temperature is approached, reaching the disordered state at the critical temperature. It should be mentioned however in the case of Cu_3Pd that as the state of complete disorder is approached and the structure is nearly cubic, the splitting of the cubic lines into their tetragonal components cannot be measured with great accuracy and so the form of the curve in this region is judged from determinations of axial ratio somewhat removed from the critical temperature. The axial ratio value of 0·9972 recorded for an alloy annealed at 460° ± 2°c would appear to indicate a slight discontinuity in this region but neighbouring values indicate a continuous variation of the degree of order with temperature.

The curves in figure 3 show that the lattice of the alloy Cu_3Pd becomes fully ordered at a temperature higher than that for the alloy CuAu. The very slow rate at which equilibrium is attained should be borne in mind in this connection; it is possible that even after the very slow cooling which specimens

of the alloy Cu_3Pd had received, they had not reached the final state of equilibrium at the lower temperatures. The alloy was judged to have been fully ordered because on prolonged annealing at these low temperatures the tetragonal parameters remained constant, but this is not a definite proof that the final state of order has been reached in an alloy which approaches equilibrium at a very slow rate. In spite of this difficulty the method used in the present investigation is the most satisfactory x-ray method of examining the ordering process in this region since the only other x-ray method available depends on the measurement of the relative intensities of the main and superlattice reflections and this could not be carried out in this region where the two groups of reflections are partly superimposed. The measurement of lattice parameters has certain advantages also over the measurement of line intensities.

Finally it should be emphasized that the measurements have been made on a unit cell which may not be the true unit cell of structure. Actually the procedure has been to follow the changes that take place in the main cubic lattice lines without paying attention to any new lines that may appear. The measurements have led to results which are consistent among themselves; they definitely show that the dimensions of the structure remain unchanged at a particular temperature after the state of equilibrium has been reached, and correspond to a definite degree of order at that temperature.

References

Jones, F. W., and Sykes, C., 1939, *J. Inst. Metals*, **65**, 419.
Wilson, A. H., 1938, *Proc. Camb. Phil. Soc.*, **34**, 81.

A Table of Chapman's Grazing Incidence Integral Ch(x, χ)

By M. V. WILKES
Director, University Mathematical Laboratory, Cambridge

MS. received 23rd November 1953

Abstract. The paper contains a table of Chapman's function

$$\mathrm{Ch}(x,\chi)=x\sin\chi\int_0^{\chi}\exp\left(x-x\sin\chi/\sin\lambda\right)\operatorname{cosec}^2\lambda\,d\lambda$$

to three decimal places for $x=50$ (50) 500 (100) 1000 and $\chi=20°$ (1°) 100° excluding values of Ch(x, χ) greater than 100.

§1. Introduction

A table is here presented of the following function which occurs in Chapman's theory of the absorption of solar radiation by an exponential atmosphere (Chapman 1931)

$$\mathrm{Ch}(x,\chi)=x\sin\chi\int_0^{\chi}\exp\left(x-x\sin\chi/\sin\lambda\right)\operatorname{cosec}^2\lambda\,d\lambda.$$

The calculations were performed on the EDSAC during the course of investigations into the use of an automatic digital computing machine for numerical quadrature.

The first programme for the EDSAC to be written for this problem was based on the use of Simpson's rule. This was later superseded by a faster programme based on the repeated application of a Gauss five point formula. In each case integration proceeded by a sequence of strips starting at the upper limit and continuing until a point was reached at which the integrand became too small to make any appreciable further contribution to the integral. The integration was then stopped, the result multiplied by the factor $x \sin \chi$, and printed. Both programmes were arranged so that a sequence of values of the integral for a given value of x and at intervals of 1° in χ was printed. It was found that the optimum strip width varied appreciably as χ increased and, in order both to economize time and to give security against error, it was found necessary to incorporate in the programme means for automatic adjustment of the strip width. A more complete discussion of this aspect of the subject will be published elsewhere.

§2. Accuracy

In the table printed by the EDSAC values of the function were given to four decimal places. Every fifth value (that is at an interval of 5° in χ) was calculated and printed both with the optimum strip width and with at least one smaller strip width. The agreement of these values provided a check against the existence of systematic errors due to truncation error in the formula used. The first two lines of the table were checked by calculating values of the function corresponding to $\chi=20°$ and $\chi=21°$ from the formula (Chapman 1931, eqn (22))

$$\mathrm{Ch}(x,\chi)=\sec\chi+\sum b_n/x^n$$

where $b_1=-\sec\chi\tan^2\chi$, $b_2=3\tan^2\chi\sec^3\chi$, etc.

χ \ x	50	100	150	200	250	300	350	400	450	500	600	700	800	900	1000
20°	1·062	1·063	1·063	1·064	1·064	1·064	1·064	1·064	1·064	1·064	1·064	1·064	1·064	1·064	1·064
21	1·068	1·070	1·070	1·070	1·070	1·071	1·071	1·071	1·071	1·071	1·071	1·071	1·071	1·071	1·071
22	1·075	1·077	1·077	1·078	1·078	1·078	1·078	1·078	1·078	1·078	1·078	1·078	1·078	1·078	1·078
23	1·083	1·084	1·085	1·085	1·086	1·086	1·086	1·086	1·086	1·086	1·086	1·086	1·086	1·086	1·086
24	1·091	1·093	1·093	1·094	1·094	1·094	1·094	1·094	1·094	1·094	1·094	1·094	1·094	1·094	1·094
25	1·099	1·101	1·102	1·102	1·102	1·103	1·103	1·103	1·103	1·103	1·103	1·103	1·103	1·103	1·103
26	1·108	1·110	1·111	1·111	1·112	1·112	1·112	1·112	1·112	1·112	1·112	1·112	1·112	1·112	1·112
27	1·117	1·120	1·120	1·121	1·121	1·121	1·122	1·122	1·122	1·122	1·122	1·122	1·122	1·122	1·122
28	1·127	1·130	1·130	1·131	1·131	1·132	1·132	1·132	1·132	1·132	1·132	1·132	1·132	1·132	1·132
29	1·137	1·140	1·141	1·142	1·142	1·142	1·142	1·143	1·143	1·143	1·143	1·143	1·143	1·143	1·143
30	1·148	1·151	1·152	1·153	1·153	1·154	1·154	1·154	1·154	1·154	1·154	1·154	1·154	1·154	1·154
31	1·159	1·163	1·164	1·165	1·165	1·165	1·165	1·166	1·166	1·166	1·166	1·166	1·166	1·166	1·166
32	1·171	1·175	1·176	1·177	1·177	1·178	1·178	1·178	1·178	1·178	1·178	1·178	1·179	1·179	1·179
33	1·183	1·188	1·189	1·190	1·190	1·191	1·191	1·191	1·191	1·192	1·192	1·192	1·192	1·192	1·192
34	1·196	1·201	1·203	1·204	1·204	1·204	1·205	1·205	1·205	1·205	1·205	1·205	1·206	1·206	1·206
35	1·210	1·215	1·217	1·218	1·218	1·219	1·219	1·219	1·220	1·220	1·220	1·220	1·220	1·220	1·220
36	1·224	1·230	1·232	1·233	1·234	1·234	1·234	1·234	1·235	1·235	1·235	1·235	1·235	1·235	1·235
37	1·239	1·245	1·248	1·249	1·249	1·250	1·250	1·250	1·251	1·251	1·251	1·251	1·251	1·251	1·251
38	1·255	1·262	1·264	1·265	1·266	1·267	1·267	1·267	1·267	1·268	1·268	1·268	1·268	1·268	1·268
39	1·271	1·279	1·281	1·283	1·283	1·284	1·284	1·285	1·285	1·285	1·286	1·286	1·286	1·286	1·286
40	1·289	1·297	1·300	1·301	1·302	1·302	1·303	1·303	1·303	1·304	1·304	1·304	1·304	1·304	1·304
41	1·307	1·316	1·319	1·320	1·321	1·322	1·322	1·323	1·323	1·323	1·323	1·324	1·324	1·324	1·324
42	1·326	1·335	1·339	1·340	1·341	1·342	1·343	1·343	1·343	1·344	1·344	1·344	1·344	1·344	1·344
43	1·346	1·356	1·360	1·362	1·363	1·364	1·364	1·364	1·365	1·365	1·365	1·366	1·366	1·366	1·366
44	1·367	1·378	1·382	1·384	1·385	1·386	1·386	1·387	1·387	1·388	1·388	1·388	1·388	1·389	1·389

χ \ x	50	100	150	200	250	300	350	400	450	500	600	700	800	900	1000
45°	1·389	1·401	1·405	1·408	1·409	1·410	1·410	1·411	1·411	1·412	1·412	1·412	1·412	1·413	1·413
46	1·412	1·425	1·430	1·432	1·434	1·435	1·435	1·436	1·436	1·436	1·437	1·437	1·438	1·438	1·438
47	1·436	1·450	1·456	1·458	1·460	1·461	1·462	1·462	1·463	1·463	1·464	1·464	1·464	1·464	1·465
48	1·462	1·477	1·483	1·486	1·487	1·488	1·489	1·490	1·490	1·491	1·492	1·492	1·492	1·492	1·493
49	1·489	1·505	1·511	1·515	1·516	1·518	1·519	1·519	1·520	1·520	1·521	1·521	1·522	1·522	1·522
50	1·517	1·535	1·542	1·545	1·547	1·549	1·550	1·550	1·551	1·551	1·552	1·553	1·553	1·553	1·554
51	1·547	1·566	1·574	1·577	1·580	1·581	1·582	1·583	1·584	1·584	1·585	1·586	1·586	1·586	1·587
52	1·578	1·600	1·607	1·612	1·614	1·616	1·617	1·618	1·618	1·619	1·620	1·620	1·621	1·621	1·622
53	1·611	1·635	1·643	1·648	1·650	1·652	1·654	1·654	1·655	1·656	1·657	1·658	1·658	1·658	1·659
54	1·646	1·672	1·681	1·686	1·689	1·691	1·692	1·694	1·694	1·695	1·696	1·697	1·697	1·698	1·698
55	1·683	1·711	1·721	1·726	1·730	1·732	1·734	1·735	1·736	1·737	1·738	1·738	1·739	1·740	1·740
56	1·722	1·752	1·764	1·770	1·773	1·776	1·777	1·779	1·780	1·781	1·782	1·783	1·783	1·784	1·784
57	1·763	1·796	1·809	1·815	1·819	1·822	1·824	1·826	1·827	1·828	1·829	1·830	1·831	1·831	1·832
58	1·806	1·843	1·857	1·864	1·868	1·872	1·874	1·875	1·877	1·878	1·879	1·880	1·881	1·882	1·882
59	1·852	1·893	1·908	1·916	1·921	1·924	1·927	1·929	1·930	1·931	1·933	1·934	1·935	1·936	1·936
60	1·901	1·946	1·963	1·972	1·977	1·981	1·983	1·986	1·987	1·988	1·990	1·992	1·993	1·993	1·994
61	1·954	2·003	2·021	2·031	2·037	2·041	2·044	2·046	2·048	2·050	2·052	2·053	2·054	2·055	2·056
62	2·009	2·063	2·084	2·095	2·101	2·106	2·109	2·112	2·114	2·115	2·118	2·120	2·121	2·122	2·122
63	2·068	2·128	2·151	2·163	2·171	2·176	2·179	2·182	2·184	2·186	2·189	2·191	2·192	2·193	2·194
64	2·131	2·197	2·223	2·237	2·245	2·251	2·255	2·258	2·260	2·263	2·266	2·268	2·269	2·271	2·272
65	2·198	2·272	2·301	2·316	2·325	2·332	2·336	2·340	2·343	2·345	2·349	2·351	2·353	2·354	2·356
66	2·270	2·352	2·384	2·402	2·412	2·420	2·425	2·429	2·432	2·435	2·439	2·441	2·443	2·445	2·446
67	2·347	2·439	2·475	2·494	2·506	2·515	2·521	2·526	2·529	2·532	2·536	2·540	2·542	2·544	2·545
68	2·430	2·533	2·573	2·595	2·609	2·618	2·625	2·631	2·635	2·638	2·643	2·647	2·650	2·652	2·653
69	2·519	2·634	2·680	2·705	2·721	2·732	2·740	2·746	2·750	2·754	2·760	2·764	2·767	2·770	2·772

χ \ x	50	100	150	200	250	300	350	400	450	500	600	700	800	900	1000
70°	2·615	2·744	2·797	2·825	2·843	2·856	2·865	2·872	2·877	2·882	2·889	2·893	2·897	2·900	2·902
71	2·718	2·864	2·924	2·957	2·978	2·992	3·003	3·011	3·017	3·023	3·030	3·036	3·040	3·044	3·046
72	2·830	2·996	3·064	3·102	3·126	3·143	3·155	3·165	3·172	3·178	3·188	3·194	3·199	3·203	3·206
73	2·952	3·140	3·219	3·263	3·291	3·310	3·325	3·336	3·345	3·352	3·363	3·370	3·376	3·381	3·385
74	3·084	3·298	3·390	3·441	3·474	3·497	3·514	3·527	3·538	3·546	3·559	3·568	3·575	3·581	3·586
75	3·228	3·473	3·580	3·640	3·679	3·706	3·726	3·742	3·754	3·765	3·780	3·791	3·800	3·807	3·812
76	3·386	3·667	3·792	3·863	3·909	3·942	3·966	3·985	4·000	4·012	4·031	4·045	4·055	4·064	4·070
77	3·558	3·883	4·030	4·115	4·170	4·210	4·239	4·262	4·280	4·295	4·318	4·335	4·348	4·358	4·366
78	3·748	4·125	4·298	4·400	4·468	4·516	4·552	4·580	4·603	4·621	4·650	4·671	4·687	4·700	4·710
79	3·958	4·396	4·604	4·727	4·810	4·869	4·914	4·949	4·977	5·000	5·036	5·062	5·083	5·099	5·112
80	4·189	4·704	4·953	5·104	5·205	5·279	5·336	5·380	5·416	5·445	5·491	5·525	5·551	5·572	5·590
81	4·447	5·052	5·355	5·541	5·668	5·762	5·833	5·890	5·936	5·974	6·034	6·079	6·113	6·141	6·164
82	4·734	5·452	5·822	6·054	6·216	6·335	6·428	6·502	6·562	6·613	6·692	6·752	6·799	6·836	6·867
83	5·055	5·912	6·368	6·662	6·870	7·026	7·148	7·246	7·327	7·395	7·503	7·585	7·650	7·702	7·745
84	5·416	6·445	7·015	7·391	7·663	7·870	8·034	8·167	8·278	8·372	8·523	8·639	8·732	8·807	8·870
85	5·824	7·068	7·788	8·277	8·638	8·918	9·144	9·330	9·486	9·620	9·838	10·008	10·144	10·257	10·352
86	6·287	7·804	8·722	9·368	9·857	10·246	10·563	10·830	11·058	11·255	11·580	11·839	12·051	12·228	12·378
87	6·816	8·679	9·867	10·735	11·412	11·963	12·425	12·819	13·162	13·463	13·970	14·384	14·730	15·024	15·277
88	7·421	9·731	11·290	12·477	13·437	14·242	14·933	15·538	16·075	16·556	17·388	18·087	18·686	19·209	19·670
89	8·119	11·009	13·084	14·743	16·139	17·351	18·426	19·392	20·272	21·080	22·523	23·784	24·905	25·914	26·830
90	8·928	12·580	15·388	17·758	19·846	21·735	23·472	25·090	26·609	28·046	30·719	33·177	35·466	37·615	39·648
91	9·873	14·535	18·401	21·868	25·091	28·148	31·086	33·935	36·717	39·446	44·789	50·030	55·211	60·362	65·505
92	10·982	17·001	22·421	27·628	32·771	37·929	43·150	48·468	53·906	59·485	71·132	83·522	96·753		
93	12·295	20·157	27·907	35·948	44·462	53·568	63·364	73·938	85·374	97·760					
94	13·858	24·257	35·575	48·371	63·032	79·920	99·422								
95	15·734	29·672	46·573	67·601	93·937										
96	18·003	36·950	62·792	98·546											
97	20·771	46·910	87·426												
98	24·175	60·809													
99	28·400	80·598													
100	33·692														

In addition the table was differenced horizontally for $\chi = 20°$, 21°, 22°, 23°. For $\chi = 90°$ the values given in the table were checked by means of the formula (Chapman 1931, eqn (17))

$$\mathrm{Ch}(x, \tfrac{1}{2}\pi) = (\tfrac{1}{2}\pi x)^{1/2}\left(1 + \frac{3}{8}\frac{1}{x} - \frac{15}{128}\frac{1}{x^2} + \frac{105}{1024}\frac{1}{x^3} - \ldots\right).$$

A check was applied to pairs of values one above and one below 90° by use of the following formula (Chapman 1953, eqn (6); there appears to be an error in the formula as there given):

$$\mathrm{Ch}(x, \chi) + \mathrm{Ch}(x, \pi - \chi) = 2 \exp(x - x \sin \chi)\, \mathrm{Ch}(x \sin \chi, \tfrac{1}{2}\pi).$$

All entries for $\chi > 90°$ were included in a check of this kind. Finally the table was checked by differencing each column on a National machine. The table was then rounded off to three decimal places. It is thought that the error should rarely exceed half a unit in the last place and should nowhere exceed one unit.

Acknowledgment

I would like to express my thanks to Miss C. M. Munford, who was responsible for checking the table.

References

Chapman, S., 1931, *Proc. Phys. Soc.*, **43**, 483; 1953, *Proc. Phys. Soc.* B, **66**, 710.

The True Contact Area Between Solids

By J. DYSON and W. HIRST
Research Laboratory, Associated Electrical Industries Ltd.,
Aldermaston Court, Aldermaston, Berkshire

MS. received 9th December 1953

Abstract. A new method for observing the points of real contact between two solid surfaces is described. A metallized glass surface is pressed into contact with the specimen, the glass surface being deformed locally at the points of contact. The opposite side of the metallized surface is examined through the glass by phase contrast microscopy, the deformations showing bright on a dark background. Deformations of a few ångströms in depth are detectable if their lateral extension is above the resolving limit of the microscope. Nominally flat, ground, polished, and lapped specimens of silver steel have been examined. With each specimen true contact was made within regions the sizes and distribution of which were determined by the general undulations of the surface and by the magnitude of the applied load. The differences lay in the distribution of the local contact areas within the general regions. With the polished specimen true contact was made over almost the whole of the general region but with the ground specimen it was confined to a few isolated areas along the grinding ridges. True contact with the lapped specimens was made at a very great number of areas each a few microns in linear dimension.

§ 1. Introduction

The cardinal fact determining the nature of the mechanism of friction and wear between nominally mating surfaces under moderate loads is that the area of true contact is in general only a small fraction of the apparent area of contact. Since the whole of the applied load is supported on this small area, these processes are essentially small in scale but severe in intensity. When the local pressures suffice to cause plastic flow, an approximate value of the total area of true contact may be obtained by dividing the applied load by the flow pressure of the softer material but relatively little is known about the sizes, shapes, and distribution of the individual areas of true contact which together comprise the total. Some information has been derived from contact resistance measurements (Bowden and Tabor 1950, Holm 1946), the principle of the method resting on the fact that the Maxwell constriction resistance due to a small circular contact region varies inversely as the radius. This method has severe limitations; the shape of the contact region has to be postulated, the individual constriction resistances due to a moderately large number of local contact areas closely grouped within a general area become swamped by the constriction resistance associated with the general area, and the measured resistance may be influenced by the presence of films of contaminant. This method, therefore, can do no more than give an indication of the number and extent of the general regions of contact and it is in principle incapable of providing information about their distribution within the apparent contact area. It is evident that a method allowing direct observation of these regions would be of considerable interest in this field.

§ 2. Experimental Methods

One method which shows directly the points of actual contact is that of Mechau (1934). This consists in placing the hypotenuse face of a right-angled prism of transparent material on the surface in question, and illuminating this face at an angle sufficiently oblique for total reflection to take place. The points of real contact are then revealed as dark areas, due to the destruction of total reflection at these points. However, due to the so-called 'optical tunnel effect', the intensity of reflected light is also reduced at points where the two surfaces approach within a small fraction of a wavelength, and so the interpretation of the image may be ambiguous. In addition, because of the obliquity of the line of sight, it is only possible to focus sharply on a narrow strip of the surface at one time. In view of these defects another method has been devised which is free from ambiguity and allows of normal viewing.

A small disc of glass with two polished faces is coated on one side with an opaque, firmly adherent film of metal. This side is placed in contact with the specimen being examined and arrangements made for applying a suitable normal load. The metallized surface is then examined through the glass by a microscope with phase contrast illumination. At the points of real contact the glass is deformed, the surface being slightly raised as seen from above. These points are then revealed, when using positive phase contrast, as bright areas. The area immediately surrounding the region of contact is deformed also, so that the true contact regions may be a little smaller than they actually appear.

The sensitivity of detection is high, for the path difference involved is approximately three times the deformation. As a result, it is possible to observe deformations of only a few ångströms in depth, provided that the lateral extent of the deformed surface is greater than the resolution limit of the microscope. The pictures shown (Plate I) were taken by this method, using a 16 mm objective; the specimens were 3/16 in. in diameter. The magnification is 69× in every case.

§ 3. Experimental Results

Some results obtained using ground silver steel are shown in figures 1, 2 and 3 (Plate I). The contacts are divided between a number of general contact regions, each containing a grouping of smaller localized contacts. Figures 1, 2 and 3 show the appearance of one of the contact regions when the total applied loads were 1·5, 2 and 3 kg, and there were about 12 general regions in all. The local contact points are, as would be expected, distributed along the ridges of the grinding marks. On increasing the load, more contacts occur within the same general contact regions whilst the length of the original ridge supports increases; simultaneously, the size of the general contact regions increases by elastic deformation of the general surface contours. The sizes of these general areas, therefore, vary reversibly with load over the load range used. The width of the ridge-like true contacts was about 3 microns and estimates of their total area showed that the local pressure would be comparable with the flow pressure.

A number of scratches on the metallized layer are visible in the photographs. These occur inevitably in use of the apparatus, but do not give rise to additional contact areas. This can be shown by their lack of change with varying load and also by examination of the strain figure in the glass by polarized light.

Figure 4 (Plate I) shows the appearance when a nearly plane metallographically polished surface of silver steel was examined under an applied load of 2 kg. Contact is made over almost the whole of an approximately circular area, the size of which agrees, to within about 10%, with that calculated by elastic theory (the calculated diameter was 0·66 mm as compared with the measured value of 0·72 mm). There is little evidence on figure 4 of localized regions of high pressure similar to those of figures 1, 2 and 3 within this general contact area. This is in accord with other evidence about the nature of polished surfaces. It has been shown (Haine and Hirst 1953) that the angles of slope of surface asperities on polished steel may be less than $\frac{1}{3}°$ and such asperities can be forced to conform to the general surface contour without applying stresses sufficiently great to cause plastic deformation. It is probable, however, that it is the local stress variations within the circle of contact which cause the whole of the contact area to be visible, for without them no detail would be visible with phase contrast microscopy. When more carefully polished specimens were used, it was in fact difficult to see the contact region.

The appearance of the steel after preparation on 'Corrosil' finishing paper was next examined (figures 5, 6, 7 (Plate II); 0·25, 1·5, 3 kg). Here again the local contact points fall within general areas, the size of which was determined by elastic deformation of the larger scale surface undulations. These local contacts are remarkably uniform in appearance and there is evidence of a residual ridge pattern; a mean width appears to be about 2 microns. The main effect of increasing the load is to increase the number of the local contacts without very much affecting their size. This is, of course, the result of the change in size of the general contact region. The effect of greatly increasing the general stress was then found. This was done by loading this specimen against the polished specimen used for figure 4 under 150 kg. On re-examination under the microscope (figure 8 (Plate II); 5 kg), the number of local contact points visible is greatly increased and they are more uniformly distributed than in figures 5, 6 and 7. The effect of the large general stress has been not so much to increase the size of the local contacts but rather to iron out small surface ripples, so increasing the number of local contacts. Examination of the polished surface after loading in this way shows that, under high general stress, occasional larger areas of true contact may, however, occur. These seem to be formed by the close clustering of smaller contacts.

§ 4. Conclusions

These experiments show that the size of the areas of true contact depend on the method of surface preparation. In so far as it is justifiable to compare results obtained with different surfaces, the present experiments using specimens prepared on Corrosil finishing paper give results in contrast with those derived by Bowden and Tabor from contact resistance measurements. They conclude that the number of true contact zones between lapped steel flats is not very large (e.g. 9 at a load of 20 kg) whilst the diameter of each zone is of the order 0·2 mm. This new method shows that the number may be enormously greater and the size very much smaller (a few microns, say). The results also have a bearing on theories now being developed to account for the frictional and wear characteristics between surfaces on which the load is supported at a plurality of asperities. In these theories (e.g. those of Archard (1953) and Lincoln (1953)), the surface is represented by a model embodying an assumed distribution of local asperities. This new method of

surface examination promises to provide evidence on the basis of which more detailed models may be formulated. In particular, the results suggest that it may sometimes be desirable to consider at least a two-scale representation of the surface irregularities, i.e. to suppose that small asperities are superposed upon longer wavelength undulations.

Acknowledgment

The authors wish to express their thanks to Dr. T. E. Allibone for permission to publish this paper.

References

Archard, J. F., 1953, *J. Appl. Phys.*, **24**, 981.
Bowden, F. P., and Tabor, D., 1950, *The Friction and Lubrication of Solids* (Oxford: University Press).
Dreyhaupt, W., 1939, *Werkstattstechnik*, **33**, 321.
Haine, M. E., and Hirst, W., 1953, *Brit. J. Appl. Phys.*, **4**, 239.
Holm, R., 1946, *Electrical Contacts* (Stockholm: Gerbers).
Lincoln, B., 1953, *Nature, Lond.*, **172**, 169.
Mechau, —, 1934, see Dreyhaupt, W., 1939, *Werkstattstechnik*, **33**, 321.

The Reduction of Choked Jet Noise*

By ALAN POWELL
Department of Aeronautical Engineering, University of Southampton

Communicated by E. J. Richards; *MS. received 2nd November* 1953

Abstract. Some methods of reducing that noise peculiar to choked jets are investigated, the mode of operation being a reduction of one or more of the factors occurring in the gain criterion $q\eta_s\eta_t\eta_d \geqslant 1$, in the notation of an earlier paper. The methods used consist of the notching of the exit edge of the jet nozzle, the introduction of cambered radial vanes and ventilation of the nozzle by means of a gauze extension. The magnitude of the noise of a jet emanating from a plain exit is shown to be susceptible to the state of the exit edge and the reflecting area presented by the nozzle body in the plane of the exit. It is suggested that the presence of initial turbulence or noise will also tend to reduce the magnitude of the sound, and this effect may be of importance in the practical case of the jets of propulsive mechanisms. The methods used influence other types of noise always present to some degree, a measured reduction in actual subsonic noise only being found in the case of the radial vanes.

§ 1. Introduction

It has been shown how the character of jet noise changes when the jet becomes choked, the noise level then increasing much more rapidly with pressure ratio than subsonically (Powell 1953 c). The possibility of jet engines as well as rocket units being operated in the choked condition seems likely. It thus becomes of importance to investigate by what means this high noise level might be reduced, if it occurs under operating conditions.

The mechanism by which the most powerful component of the noise of such a jet has been examined (Powell 1951, 1953 a, 1953 c) and, although qualitative results only are so far available, a suitable line of approach is apparent. The condition for the sound-producing motion to be maintained has been given as $q\eta_s\eta_t\eta_d \geqslant 1$ in the previous notation, i.e. q is the amplification of the stream disturbance as it passes from the orifice, where it is created by the passing sound wave with an 'efficiency' η_d to the effective source where interaction with the shock waves produces acoustic energy at an efficiency η_s. This sound is transmitted back to the orifice with efficiency η_t. The foregoing criterion is applicable whether the motion is periodic (producing a 'screech') or irregular, although in the latter case the terms may become more complex—for example η_t will no longer be the simply determined directionality previous considered.

Therefore, if the jet conditions can be so modified that the product $q\eta_s\eta_t\eta_d$ can be reduced, then the amplitude of the motion will also be reduced, it being eliminated altogether if the criterion cannot be met at all. Conversely, the reasons for observed differences in noise level for various conditions may be sought for in variations of that product. The importance of certain of the

* Formerly A.R.C. 15624 F.M. 1858 a, December 1952. This paper is a sequel to "The Mechanism of Choked Jet Noise", *Proc. Phys. Soc.* B, 1953, **66**, 1039.

individual factors had been suspected from the start (these being q and η_a), but the fundamentally important nature of the *product* of factors was not realized until some time later. Some of the properties of such a criterion have been deduced (Powell 1953 b, 1953 c), and it has been used as a guiding principle in the later stages of the work.

It should be noted that the experiments in the main are of an exploratory nature to investigate the scope of various schemes of noise reduction: it is not unlikely that a systematic variation of the parameters involved would result in improved results for the total noise level, since in most cases the actual reduction is the net difference between the reduction of the characteristic 'screech' and other changes, usually increases, introduced by the device itself.

Subsonically conditions are much less amenable to analysis, except in the special case when periodicities are present (for example, due to breakaway, Powell 1951), and it will be recalled that experiments on the influence of velocity profile produced rather disappointing results so far as economic noise reductions were concerned (Powell 1952). For this reason the measurements of noise intensity have been extended into the subsonic regime in all cases as a purely *ad hoc* investigation in that region. In view of the great importance of subsonic jet noise today a special section is devoted to this aspect. It must be remembered that the subsonic type of noise, i.e. that due to turbulent mixing, is always present in a choked or supersonic jet although not normally the dominant noise.

§ 2. Apparatus

The noise measurements were made for an air jet of two inches diameter exhausting into the atmosphere, within the confines of a laboratory. The air supply for the jet was taken from a large reservoir especially provided for the purpose, the flow being controlled by manually operated gate valves in conjunction with a pressure reduction valve. Silencers were placed in the six-inch diameter main after these control valves in order to reduce as far as possible any valve noise, the pressure of the flow after these and just prior to the contraction to the nozzle exit being measured. The pressure p_j' so measured was very nearly the dynamic head of the jet, termed 'jet pressure', and is a useful parameter. The pressure ratio of the jet for all practical purposes is $(p_a+p_j')/p_a$, where p_a is the atmospheric pressure. The temperature of the flow, also just prior to the contraction, was measured, it being found that the difference from that of the ambient air was insignificant.

Except for the first nozzle, designated A, which had an ordinary bell mouth contraction, the contraction of the nozzle was designed so as to reduce the possibilities of breakaway of the flow from its walls. The latter nozzle has been designated B, and several changes were made at the exit, the body being designed so that various appendages could be attached. The various exit conditions are described in the section dealing with their influence on the noise of the jet. A cone of absorbent material was placed on the body of nozzle B in order to reduce reflections to a minimum (see figure 1).

A Standard Telephones and Cables objective noise meter Type 74100 A was used in conjunction with a Bruël Kjaer high speed level recorder for the noise measurements. The latter replaced the meter normally used, and enabled many more readings to be obtained during the limited duration of the tests, also

enabling a mechanically averaged value of the fluctuating noise level to be obtained. This proved most valuable, apart from a permanent record on waxed paper being produced, since it could be inspected as fast as it was made. Owing to the nature of the averaging of the recorder, and the response of the moving-coil microphone and associated amplifier, the 'total' noise measurements have been subjected to an effective response. This is given in table 1, estimated for measurements of noise like that of subsonic jet flow; measurements of noises approaching pure notes may suffer a further 4 decibels decrease.

Table 1

Filter bandwidth (c/s)	37·5	50	75	100	150	200	300	400
	75	100	150	200	300	400	600	800
Effective response (dB)	−6·5	−5·0	−4·0	−3·0	−1·0	−0·5	0	0
Filter bandwidth (c/s)	600	800	1200	1600	2400	3200	4800	6400
	1200	1600	2400	3200	4800	6400	9600	12800
Effective response (dB)	−0·5	−1·5	−3·5	−4·0	−5·0	−4·0	−5·0	−6·0

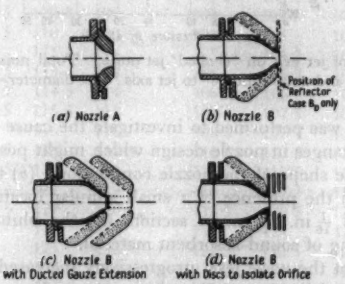

(a) Nozzle A
(b) Nozzle B
Position of Reflector Case B_D only
(c) Nozzle B with Ducted Gauze Extension
(d) Nozzle B with Discs to Isolate Orifice

Figure 1. Various nozzle configurations (exit diameter=2 inches).

Owing to the fact that the noise measurements were made within the confines of a laboratory, certain systematic errors due to reflected noise will be present. The effect of these on the comparison of various noise levels will be small. Subsonically the errors cannot be greater than about +3 dB at 90° and +2 dB at 30° to the jet direction, both positions being four feet from the nozzle exit. These values are unlikely to be exceeded in other cases. These are the standard measuring stations.

Photographs of the flow were taken using a two-mirror Toepler–Schlieren system, with the illumination provided by a spark between magnesium electrodes, the effective duration being about a microsecond. The nozzle and models used in this work were half-scale replicas of those used in the noise measurement experiments.

§ 3. Influence of Exit Conditions

The first indication that the efficiency of the noise production above choking was susceptible to edge conditions became apparent at an early stage in the work when the temporary nozzle A was replaced by model B_R (Powell 1951). The

suffix R indicates that the exit edge was rough, due to a slight crumbling of the material at a few points during machining, the maximum size of the cavities produced being less than 0·03 inch. As can be seen from table 2 and from figure 2, noise reductions of as much as $7\frac{1}{2}$ decibels have occurred above choking, whilst subsonically the change is insignificant.

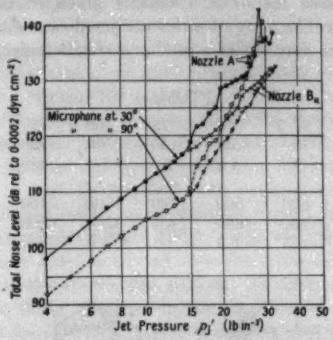


Figure 2. Influence of jet exit on 'choked' jet noise. Total noise levels, measured 4 ft from exit and at 30° and 90° to jet axis. Jet diameter = 2 inches.

A series of tests was performed to investigate the cause of this difference in noise level. The changes in nozzle design which might possibly be responsible are (figure 1) (*a*) the shape of the nozzle contractions, (*b*) the condition of the actual exit edge, (*c*) the presence of a small annular locating ring continuous with the bore, and $\frac{1}{16}$ in. $\times \frac{1}{16}$ in. in section, (*d*) the blunt face of nozzle A replaced by a backing of sound-absorbent material.

It was found that the noise level progressively increased towards its former nozzle A value as the following modifications were made to nozzle B_R:

(i) The addition of a plain ring lengthening the bore by $\frac{1}{2}$ in. and the filling in of the pitting at the original exit edge, there being a location annulus and the exit edge of good finish. This is case B_P.

(ii) The removal of the annulus. Case B_T.

(iii) The introduction of a reflector in the plane of the exit (figure 1(*b*)). Case B_D.

The noise levels obtained under these conditions are shown in table 2 with the original values. The influence of the actual condition of the exit edge has already been commented upon (Powell 1951), and the suggestion that damping of the stream disturbances is enhanced by the turbulence introduced by 'wakes' streaming from the irregularities at the edge appears valid.

The change (iii) was made on the grounds that if the stream disturbances are sound-induced at the orifice, then a 'rigid' reflector there would increase the local sound pressures and so amplify the motion. The fact that the sound intensity is found to be susceptible to such a reflector hardly interfering with the stream is considered as additional evidence in favour of the mechanism suggested in previous papers. It is interesting to notice that over one range the sound levels have increased beyond their former values, the conditions for the particular

disturbance arrangement apparently being especially susceptible to an increase in the overall 'efficiency' of the cycle.

The effects described may be responsible for the decrease in noise level by 3 decibels recorded by other workers (Westley and Lilley 1952) when the external contours of a one-inch diameter jet were altered so as to produce a sharp, acute-angled edge to the nozzle, instead of a blunt nozzle end.

Table 2. Influence of Nozzle and Edge Conditions

Total noise levels in decibels above 0·0002 dyn/cm^{-2}.
Two-inch diameter jet, measurements 4 ft from exit, at 30° and 90° to direction of jet.

Jet pressure p_j' (lb in^{-2})	At 30° station Nozzle				At 90° station Nozzle				
	A	B_T	B_P	B_R	A	B_D	B_T	B_P	B_R‡
4	98	97·5	98·5	98·5	91·8	92	91·5	92·5	92·5
5	101·5	100·9	102	102	95	95·2	94·5	95·3	96
6	104·5	104	104·8	105	97·8	98·5	97·2	98·2	99
7	107	106·5	106·8	107·5	100·2	100·1	99·5	100·2	101
8	108·5	107·5	108·5	109	102	101·8	101	101·7	103
9	110·5	109·5	110	110·5	103·5	103	102·5	103·4	104·5
10	111·8	111	111·5	112	105	104·3	104·2	104·6	106
11	113	112·5	113	113	106	105·7	105·5	105·5	106·5
12	114	113·1	114·5	114	106·5	106·5	106·5	106·8	107·5
13	115·2	114·5	115·8	115	107·5	107·5	107·2	107·8	108
14	116·5	116	116	116	108·5	107·7	108	108·2	109
15	118·8	117·5	118	117·5	110·5	110·5	110	109	109·5
16	121·5	119·5	121	118	114·5	115·5	112	110·5	111
17	122	122·4	122	119·5	116	117·5	115	112·5	113·5
18	124	123·5	122	120	118·5	120·5	120	114	115·5
19	125	124	122	121	119	121	121·8	115·5	117
20	128·5	125	123	122	121·5	122·5	123	118	118·5
21	129	*	125	123	123·5	123·5	123·5	119·5	119·5
22	129·5	128	126·5	124	125·5	124·5	124·5	121	121
23	130	129	128†	125	127	126·5	127	123·5*	122·5
24	130·5	130	129·5	127	128·5	127·5	128	125·7	124
25	131	131	130	128	130·5	128·5	129	126·8	125
26	132·5	132	130	128·5	133·5	>140	129·8	128	125·5
27	134·5	132·7	130	129	135·5	>140	130·5	128·7	127·5
28	*	133·2	131	129·5	136·5	>140	131	129·5	128·5
29	137	134	131·5	130·5	*	>140	132	129·7	129·5
30	137·5	134†	134	131	137·5	133·5	133	130	130
31	136·5	134	134	132	138·5	—	133·5	131	131
32	138·5	134·5	136	132	137·5	—	134	132	131·5
33	—	135	—	—	138·5	—	134·2	133·5†	132·5
34	—	137	—	—	140·5	—	135	134	—
*Max.	142·5	132	128	—	140·5	—	126	128	—
†Max.	—	138	—	—	—	—	—	140	—

‡ Readings have possible error of +1 dB throughout.
Maxima occur at or near indicated positions.

In general, unless the amplitude of the motion is effectively stabilized, say by the introduction of a large degree of turbulence, the noise levels above choking appear highly susceptible to the conditions pertaining at the time of test. It has

been found that they may vary by one or two decibels on repeat tests conducted within a short time of each other, whereas the subsonic values remain unaltered. For this reason the noise levels in table 2 should be considered as illustrating a general trend, rather than as specific changes associated with each modification.

§ 4. Notching of the Orifice

It has been shown how a slight roughness of the exit edge creates sufficient turbulence in the boundary layer to result in a measurable decrease in noise level above choking. This suggests that the notching or nicking of the exit edge may introduce additional turbulence into the stream, effective in reducing q and so diminishing the sound amplitude. Offsetting this will be noise generated by the turbulence itself and, perhaps, by its traversing the shock system: the net sound reduction can be found experimentally.

The notches were made in short rings which were attached to the exit of the two-inch diameter jet. One of these rings is shown in figure 3 (*a*)*. The form of notches first tested was flat-sided Vee, $\frac{1}{10}$ in. deep, tapering to zero half an inch from the exit, the 90° included angle being maintained. Four notches were introduced, producing either a large reduction in noise level or no reduction at all. This inconsistency was attributed to occasional breakaway occurring at the angular shoulders of the V-notches, and these were subsequently rounded off so that no flats remained and the transitions were quite smooth. Noise recordings were made with four, eight and finally sixteen symmetrically placed notches.

Influence on sound level.

The measured noise levels showed a progressive reduction with increasing numbers of notches. The levels at both measuring stations are shown in figure 4 for the least noisy case, i.e. for sixteen notches.

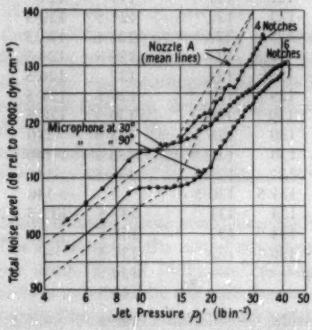


Figure 4. Influence of notching jet orifice. Total noise levels at 4 ft from exit at 30° and 90° to axis. Jet diameter=2 inches. Notches: round-shouldered Vee with (notch depth)/(jet diameter)=1/20.

A second series of nozzles was tested; here up to twelve notches of similar geometry, but twice the size, were used. The noise levels above choking were

* Figures 3, 5, 6, 7 ,14, 15, 16, 17, 19, 20 are printed as Plates I–III at the end of this issue.

essentially the same as for the more numerous smaller notches, but the increase in subsonic noise level (see figure 4) was noticeably larger.

Modification to the flow.

Schlieren photographs clearly show the increased spreading of the jet due to the notches: in figure 5 a single notch was present at the lower edge of the orifice. The flow can be clearly seen as it leaves the notch parallel to the diverging main flow at the exit, and appears to form the jet boundary in this direction, the consequent spread being greatly enhanced on that side of the jet. The very high frequency sound appears to have been reduced, it being clearly visible on the notchless side of the stream.

The flows from a plain exit, and from one with sixteen notches, can be compared in figures 6 and 7, for which the Schlieren cut-off was parallel to the stream. The 'ridges' in the flow from the notches are prominent, although the later spread of turbulence is not so clearly seen. There is a suggestion of much very high frequency *noise* emanating from the modified stream, possibly partly from the main shock position, although the source cannot be located with any confidence. In the cases shown, the normal bridge or Mach shock had appeared, eliminating the apex of the conical shock, and the changed conditions behind this formed a clearly defined cylinder. It will be noted that, although the spreading of the flow has been greatly increased, the cellular structure of the flow, with its shocks, is hardly affected (figure 5).

Refraction of 'shock' noise.

A certain amount of noise may be generated by the turbulence traversing the shock system, which will be termed 'shock' noise. It is important to note that it is a *noise*, having no more definition in frequency than the general turbulence responsible. Previous considerations have dealt with the case when the disturbances involved the jet boundary, and the sources of sound were shown to be effectively situated and stationary in the external fluid (Powell 1953 c). The case is now rather different, for the noise will be created within the moving stream and will therefore undergo refraction due to velocity and temperature gradients before emerging into the external atmosphere.

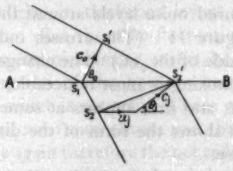


Figure 8.

The results of refraction can be examined qualitatively in the following greatly simplified manner. First of all consider a jet of velocity u_j bounded by two plane velocity discontinuities, one of which is AB (figure 8). The points s_1, s_2 on a plane sound wave moving at angle θ_j and velocity c_j relative to the stream will move to s_1' and s_2' unit time later, the wave front having emerged from the stream at an angle θ_0. If the plane wave be considered as part of one of large

curvature, the radiation in a given direction will increase in the ratio $R = d\theta_j/d\theta_0$, due to refraction across the boundary. Since $c_0 \sec\theta_0 = u_j + c_j \sec\theta_j$, then

$$R = \frac{c_0}{c_j}\frac{\tan\theta_0 \sec\theta_0}{z(z^2-1)^{1/2}} \quad \text{where} \quad z = \frac{c_0}{c_j}\sec\theta_0 - \frac{u_j}{c_j}.$$

Reflections resulting from impedance mismatch across the boundary have been ignored, since all the energy would be ultimately transmitted at the same angle. The following approximate values may be used: $u_j/c_j = 1$, $c_0/c_j = (\gamma+1)^{1/2}/\sqrt{2}$. The result is shown in figure 9. An infinite peak occurs at $\theta_0 = 56{\cdot}75°$, no sound being transmitted at angles less than this. An angle of total internal reflection occurs (at $\theta_j = 118{\cdot}5°$) but, like other cases for θ_j, $\theta_0 > 90°$, this has no meaning unless the sound is assumed to be effectively generated in a subsonic region, e.g. at the end of a decaying shock with diffraction.

The existence of an infinite peak (at $\theta_0 = 56{\cdot}75°$) is a consequence of the disregard of the wave equation. In practice lateral spreading of the wave would take place (diffraction), and in a circular jet further distribution would occur through reflection, and scattering due to its passage through turbulence. These effects might be allowed for by applying a distribution factor of the form $\exp(-k\theta_0)$, applied to summations over, say, 10° intervals. This has been done in figure 9 for $k = 1/20$, the peak having shifted its maximum to about 60°. Notice that these values, and the form of the factor chosen, are arbitrary.

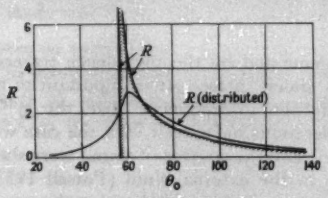


Figure 9. Directionality due to refraction.

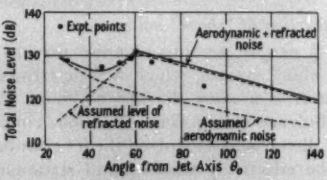


Figure 10. Comparison of estimated distribution with experiment.

A series of tests was performed to find whether the peak so predicted is in fact the case. The measured noise levels around the jet over a wide range of pressure are shown in figure 11. (The crosses indicate typical check points obtained on the opposite side of the jet.) The change in directional distribution which takes place above choking is most noticeable, being quite different from that of the subsonic noise, and first appears at some pressure greater than the critical. The broken line shows the form of the distribution at $p_j' = 40$ when corrected for reflected noise.

A maximum is found to occur at the predicted angle of 60°, the peak being in fact quite sharp; the closer measuring points were some three inches apart, and the microphone head was two inches in diameter. A curve can be constructed by taking a suitable level of noise, with the directionality of the subsonic noise due to turbulent mixing, and adding refracted 'shock' noise to produce the peak. This has been done in figure 10. The general agreement is reasonable, qualitatively. Better agreement might be obtained by adjustment of the distribution factor, or varying the 'shock' noise distribution before refraction,

but there is no point in doing this until such steps can in themselves be justified, the analysis itself being of an exploratory nature.

As can be seen from figure 11, the tests disclosed another peak around $\theta_0 = 100°$. This is thought to be residual noise of the type which it was desired to suppress for which a peak would be expected near that position. It should be noted that the noise distributions were done with notches present, modifying the flow over the whole range of pressure ratio. Above choking, the flow appears to be stabilized to some extent, since noise measurements are repeatable (to within instrument accuracy), whereas with a plain clean exit considerable variations might occur.

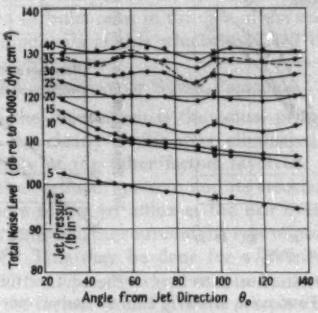


Figure 11. Measured noise distribution around jet at 4 ft radius. Broken curve for $p_1' = 40$ corrected for reflected noise.

§ 5. Introduction of Radial Vanes

The notching of the orifice resulted in a decrease in the choked noise by virtue of the increased boundary layer thickness, but the associated turbulence gave rise to increased subsonic noise. If radial vanes developing a lift force were introduced (see figure 3(*b*)), trailing vortices might extend along the boundary of the jet. These would modify the boundaries and encourage momentum transfer, resulting in thickened layers of vorticity, and so reduce the disturbance instability. But the time-fluctuations in velocity at any point associated with this momentum transfer might be expected to be decreased, because the rotational motion of the vortices carrying the axial momentum changes has a mean value which is not zero. There will of course be interchanges between the vortex-carried axial momentum and the surrounding flow, and this may produce noise. Once again therefore the net reduction in total noise depends upon the difference between that eliminated and that introduced simultaneously. In addition the vanes may influence the structure of the flow as a whole.

The form of the vanes was as shown in figure 3(*b*). The vanes were of 16 s.w.g. copper, suitably profiled and having a length parallel to the flow of half an inch. The penetration into the flow from the boundary varied up to three-quarters of the exit radius, which was one inch as before. Camber was introduced to the vanes by displacing the trailing edges which were proud of the exit; the amount of camber was varied and also its sense with regard to neighbouring blades.

Influence on sound level.

Noise measurements were first made for uncambered vanes, and these were found to be almost identical with those for the sharp-edged exit B_P alone. The wake associated with the profile drag therefore has no measurable effect.

Six cambered vanes were more effective than three, and camber resulting from a $\frac{1}{8}$-in. trailing edge displacement was more effective than for $\frac{1}{16}$ in., but it was immaterial whether the vanes were cambered all in the same sense or alternately in opposite senses.

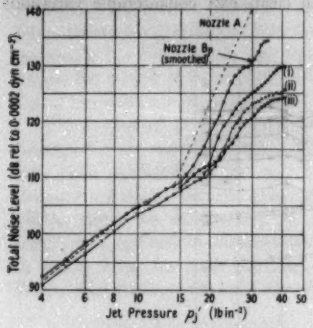


Figure 12. Variation of noise level at 90° with vane depth. Total noise levels at 90° 4 ft from jet exit. Addition of 6 cambered radial vanes of depth (i) $\frac{1}{8}$ in. (ii) $\frac{1}{4}$ in. (iii) $\frac{1}{2}$ in. to nozzle B_P of 2 inches diameter.

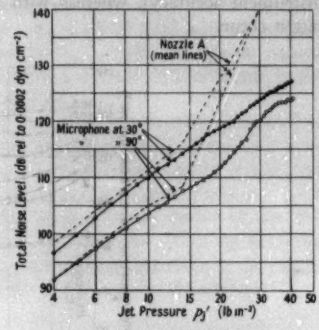


Figure 13. Influence of six cambered radial vanes. Total noise levels at 4 ft from exit and at 30° and 90° to jet axis. Six cambered radial vanes of 2 in. depth. Nozzle diameter 2 inches.

The progressive decrease in noise level with the depth which the vanes penetrate the flow further from its boundary is illustrated in figure 12, which shows the noise levels at the 90° station. Six vanes were present, having a trailing edge displacement of $\frac{1}{8}$ in. (figure 3 (*b*)). It was found that there was no further decrease in noise level when the penetration exceeded half the orifice radius. The levels at both measuring stations are shown in figure 13 for that vane size, a comparison with the original nozzle A again being made.

Aurally, the 'screeching' did not appear at all; there was however a very high-pitched hiss emanating from close to the vanes, the subjective intensity being much lower than the other noise.

Schlieren photographs of the flow.

Two typical photographs are shown in figures 14 and 15 for a single vane present, in two different planes. The interpretations from the photographs have been aided by exploration of the flow with a short fine silk streamer attached to a fine wire. The most important features disclosed are the tip vortices (compare with figure 7), the inner one being compact and almost straight, but the outer one having a large 'downwash' and rapidly expanding into the jet boundary, the spread being seen particularly in figure 15. In addition, several shock waves arise, but their geometry is difficult to determine. Two strong shocks arise from the undersurface of the vane (clearly shown with a vertical cut-off, but only very

faintly with it horizontal, as in figure 14), and a pair extending towards the centre of the flow appearing from the trailing edge are also apparent (figure 15).

The presence of six vanes modifies the flow in a marked fashion, figures 16 and 17 (cut-offs respectively normal and parallel to the stream). A very complex structure of shock waves exists near the orifice, the flow rapidly expanding and, although the gradients characteristic of the usual cellular pattern are present, the structure is weakened and very diffuse. Thus, while the notches affect only q to any great extent, the vanes reduce both q and η_s. High-frequency sound radiation is present in both cases, part of this moving away at about 45° to the stream as observed from the plain jet, the boundary layer still appearing initially laminar, but other radiation normal to the stream is also present, apparently generated by the induced turbulence as in the case of the notches. It is perhaps the lower frequency components of this which produced the audible 'hiss'.

§ 6. Removal of Super-Pressure

An essential part of the mechanism is the actual production of the sound output, associated with η_s: clearly if this were eliminated there would be no output quite independently of the other factors involved. Since this depends upon the formation of the 'cellular' pattern and its associated shocks it follows that if the excess pressure of the jet efflux at the exit over the ambient value, i.e. the super-pressure, is removed, the source of the type of noise we are considering would also be removed. This may be done for a given initial pressure ratio by expanding the flow either by a divergent exit or by direct ventilation of the nozzle walls. Whereas the former would give the correct expansion for a unique pressure ratio, the latter method would be applicable over a much wider range, though being perhaps rather less efficient. Such ventilation might be achieved by introducing slots into the nozzle wall, but a neat method is to add a cylindrical gauze extension,* and this was used. Its length was two diameters, and was formed of 40-mesh 0·008 in. diameter wire (figure 3 (*c*)).

Influence on noise levels.

Tests with the gauze sealed gave choked noise levels greater than for the plain nozzle with a rough exit edge B_R but less than for it when sharp-edged, B_P, the gauze mesh having been blocked in by solder to form a fairly smooth surface, but less smooth than a machined ring. This result is as would be expected.

Preliminary tests showed that the outflow through the open gauze, even for low subsonic exit velocities, created a considerable noise, audible as a 'hiss', and this predominated at low exit velocities (see Powell, 1951 where this section was first reported). This was partially overcome by ducting this secondary flow away, the duct sides being packed with sound-absorbent material (fibre glass) as indicated in figure 1 (*c*), and although this much improved conditions an appreciable amount of this noise still remained, appearing to emanate from the jet orifice. The increase in subsonic total noise level, at least partly due to this, can be seen in figure 18, where the results are presented. Above choking considerable reductions are found, the level at 30° increasing as p_j' to the power 3·0 over the whole range investigated (and in fact still applies at $p_j' = 0{\cdot}25$ lb in^{-2}, Powell 1951). At 90° the levels are again reduced and, in contrast to the 30° levels, a marked humping is present, this having become accentuated after

* Such a method was suggested by Professor A. D. Young.

the ducting was introduced, which presumably reduced the ventilation efficiency. This suggests that the fall and rise of the noise level are the result of incomplete ventilation and that they may be caused by a directional minimum and maximum swinging through 90°, this explanation being considered more likely than that there are changes in the total emitted energy.

Modifications to the flow.

Several changes result from the outflow through the gauze, leaving aside the noise it may create and the weakening of the shock-wave formations: the mass flow is reduced, the jet pressure reduced, the stream accelerated, and boundary layer suction would result in a thin boundary layer at the exit. In figures 19 and 20 the flows with the gauze sealed and open are compared for the same pressure ratio. In the latter case the flow is supersonic although a slight excess pressure appears to remain.

If this expansion were considered complete, then the Mach number would rise to about 1·415. Plotting the 30° noise level various parameters gives the following results: over the whole range (subsonic and supersonic) noise energy depends on the jet pressure p_j' to power 3·0; on an exit velocity basis this is equivalent to a variation between $V^{5·75}$ at the lowest velocities to $V^{10·5}$ at the highest. It should be remembered that, in addition to the jet noise due to turbulent mixing, there is additional noise due to the outflow through the gauze which is important subsonically, and some shock noise may arise in the supersonic case. Nevertheless it is useful to bear in mind Lighthill's (1952) theoretical estimate of V^8 for *low* Mach numbers in considering the upper end of the range.

§ 7. Isolation of the Orifice

The effect of the generated sound at the orifice is an important stage in the self-maintained motion producing the 'screeching' effects. If this coupling between the sound and the stream could be weakened or eliminated, then the periodic motion would follow suit. An attempt to achieve this was made by introducing the discs of acoustic tile as sketched in figure 1(*d*). Several minor modifications were made in an attempt to eliminate the coupling without excessive interference with the stream, there being an appreciable inflow between the discs. The aim of eliminating that particular type of noise was achieved, but there was hardly any reduction in total noise levels owing to the appearance of much *noise* introduced by the device itself; it is thought that this is largely due to the inflow breaking away from the inner edges of the discs and so introducing an irregular turbulence into the stream which results in the additional noise, possibly at least in part due to interaction with shock waves. These tests were therefore largely inconclusive in that this turbulence might also be responsible for the elimination of the periodic motion, in the same way as for the notches, for example. Clearly from the point of view of overall noise reduction nothing is to be gained, although it may be possible to modify the arrangement so as to reduce the troublesome effects associated with the inflow.

§ 8. Results in the Subsonic Regime

The various schemes described above will be dealt with in turn.

(i) *Notching of the orifice.* It will be recalled that this had the effect of introducing turbulence into the stream near the orifice, and produced additional

noise due to the fluctuations across the shear layer. It can be seen that the increase is velocity- (or jet pressure-) dependent, the increase being fairly small at low jet speeds and rising to a maximum increase of about 3 decibels at $p_j = 10$ lb in^{-2} (Mach number $M = 0{\cdot}9$), but thereafter decreases so that effectively the total noise remains constant while the jet speed is increased to sonic where the noise level is the same as in the notchless case.

The latter is an important observation since it may well be the case that such practical jets as this device might be applied to will either be idling, when the noise produced is unimportant, or operating from the critical pressure upwards, where the notches produce an increasing reduction in noise level.

(ii) *Radial vanes.* These vanes (figure 3 (*b*)) were fashioned so as to enhance the mixing without introducing undue turbulent fluctuations. There was a decrease in subsonic noise level of about 2 decibels (figure 13), so that this method results in a noise reduction over the whole range of jet pressure. No thrust measurements have been made, although the change in mass flow was too small to be measured.

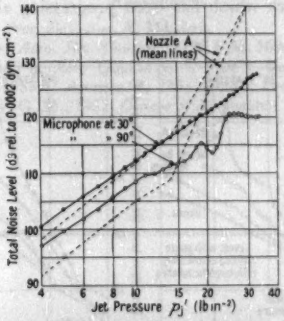


Figure 18. Influence of ventilation of nozzle. Total noise levels at 4 ft from exit and at 30° and 90° to jet axis. Ventilation by cylindrical gauze extension two diameters long.

(iii) *Rotation of flow.* The radial vanes resulted in a decrease in overall noise level, and much of this arises from an appreciable distance downstream. It was thought that it might be the general rotation imposed upon the flow, due to the 'downwash' behind the cambered vanes. To investigate this vanes were introduced into the approach pipe prior to the nozzle contraction, and resulted in a rotation of the stream of about 5°. There was no appreciable change in subsonic noise levels, and none above choking except over a certain range where there was a large increase. This range was identical to that over which the increase was noted when the reflector was placed in the plane of the exit (table 2).

The inference is that vortices lying along the edge of the jet might cause a small reduction, although a general rotation does not. This aspect may well be worth further examination.

(iv) *Gauze extension.* The outflow through the gauze sets up an additional noise, while there is no reason to suppose that any noise reduction occurs elsewhere. There is thus an overall increase in the subsonic noise level (figure 18). The same is true of the arrangement of radial discs (figure 1 (*d*)), the additional noise in that case arising from the interference of the inflow with the main flow.

§ 9. Concluding Remarks

It has been shown in the earlier paper (Powell 1953 c) how the noise of a choked jet is made up of several parts:

(i) That due to turbulent mixing, as in a subsonic jet. The noise created by the collapse of the initially laminar boundary layer may also be included in this section.

(ii) That due to the passage of turbulence across the shock waves in the flow. This may be divided into two parts, (*a*) irregular turbulence, (*b*) a periodic motion which is self-maintained.

In the types of jet investigated in this work, the last of the above groups (ii (*b*)) was by far the largest above choking. It is shown in this paper how this noise can be reduced and practically eliminated by introducing additional turbulence into the stream, weakening the shock waves and, to a lesser extent, by reducing the size of the reflecting surface formed by the end of the nozzle body. The mechanism by which the reductions come about can be explained in terms of the theory of this choked jet noise as suggested in the introduction, as may similar reductions in other work which has been progressing simultaneously (Westley and Lilley 1952).

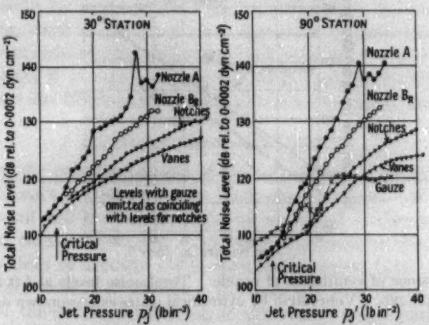


Figure 21. Comparative noise levels for various conditions.

Although the radial vanes result in the largest noise reduction, about 20 decibels, the notches clearly present the most practical method although the reduction is a few decibels less, the associated increase in noise level over part of the subsonic regime probably being out of the working range. The noise levels obtained by the various methods is shown in figure 21, enabling a comparison of the effectiveness of the methods to be made. In all cases, however, the effective total noise reduction as measured depends also upon how much other noise is introduced by the particular device in question, and this may apply to the subsonic case as well. It is useful to note therefore that of the devices tested the radial vanes is the only case where a measured reduction of the subsonic noise occurs, although even here the frequency coverage was incomplete.

Another point of considerable practical interest is the effect of initial turbulence, or noise, in the jet stream. The importance of this arises because of the dependence of the sound-generating motion on the degree of instability of

the disturbances, which is of course dependent upon the turbulence and spread of the flow, initial noise possibly initiating other disturbances introducing further damping. Thus this type of noise might appear in varying degrees in practice, and if the initial conditions are turbulent and noisy enough (for example, perhaps in the efflux of a turbojet) may not appear at all. It has also been shown how the conditions at the exit edge, i.e. its roughness and shape, influence the magnitude of the sound. Further, the higher exit velocity of many propulsive mechanisms (due to elevated temperature) will produce considerably more noise due to turbulent mixing, so that the characteristic choked jet noise may form a less important contribution to the total noise. For these reasons the total amount of noise reduction attainable is obviously dependent upon the amount of this type of noise peculiar to the choked jet which is initially present. The efficacy of the methods will be really known when they are applied to full-scale jets, but it is satisfying to know that if this type of noise does arise, then the problem of its reduction can be tackled with a good chance of success.

References

LIGHTHILL, M. J., 1952, *Proc. Roy. Soc.* A, **211**, 546.

POWELL, A., 1951, *Report Aero. Res. Council* 14726 F.M. 1694; 1952, A.R.C. 15473 F.M. 1837, to be published in *Aero Quarterly*, **4**, Part 4, February 1954; 1953 a, *Aero Quarterly*, **4**, 103; 1953 b, *Acustica*, **3**, 233; 1953 c, *Proc. Phys. Soc.* B, **66**, 1039.

WESTLER, R., and LILLEY, G. M., 1952, *College of Aeronautics Report* No. 53.

Observations on a Form of Breakdown in Germanium Diodes

By A. Lempicki and C. Wood
Electronic Tubes Ltd., High Wycombe, Bucks.

MS. received 1*st January* 1954

Abstract. The (I, V) characteristic of germanium diodes is shown to depend markedly upon the wave forms used to plot it by oscilloscopic means. When the time variation of current is studied by the use of microsecond pulses a breakdown first observed by Billig occurs. The critical barrier temperature at which this breakdown occurs and the barrier radius are obtained by postulating a simple model for the contact. The possibility of the breakdown being due to thermal instability is discussed.

§ 1. Introduction

For some time the phenomenon of breakdown in germanium point contact rectifiers has been recognized as due to self-heating effects and not to purely electrical properties of the contact. When diodes are investigated under conditions in which thermal effects are eliminated the current is a monotonic single-valued function of the applied voltage. No very sudden increase of current takes place for voltages several times larger than the turnover voltage. An extensive theoretical and experimental analysis of the inverse characteristic at high fields has been carried out by Simpson and Armstrong (1953).

In contrast to this stands the well-known shape of the (I, V) curves obtained by oscilloscopic means and exhibiting voltage turnover and a region of zero differential resistance. Although the phenomenon of turnover does not lend itself to detailed analysis there is little doubt that it is due to self-heating effects (Hunter 1951, Tillman and Henderson 1953).

A demonstration of the fact that the shape of the (I, V) characteristic depends upon the time the current is allowed to pass, and hence its thermal state, can be given by displaying the curves on an oscilloscope using different wave forms. Figure 1 gives characteristics of two diodes (Nos. 1 and 2) traced by using a sine wave and short pulses. The difference in appearance of these curves is very striking. The two diodes were representative of two classes of commercial diodes which were available for this investigation.

The turnover voltage as observed and measured using the a.c. display is an ill-defined quantity. Its value depends very markedly upon whether a voltage symmetrical about zero or unsymmetrical is used, and whether or not a current-limiting resistor is in series with the diode. The duration of an experiment is also very important. In some cases (such as figure 1 (*a*)) we can obtain a turnover voltage of about 160 volts by rapidly increasing the sweeping voltage from zero. Once the turnover is reached the whole curve will start to shrink slowly and the turnover voltage may decrease by as much as 50%. For this reason it seems not possible to attach much importance to measurements of turnover voltage and current either by a.c. or d.c.

A different manifestation of breakdown is obtained when we observe the variation of the current flowing through the diode with time. An experiment of this kind was performed by Billig (1952) but no detailed investigation has been carried out. Billig observed that when a constant voltage pulse is applied to a diode the current pulse is first flat-topped but shoots up in a characteristic fashion

Figure 1. (*a*) Diode No. 1; inverse characteristic obtained using 50 c/s rectified sinewave; (*b*) same as (*a*) for diode No. 2; (*c*) current and voltage pulse (duration 3 μsec, repetition frequency about 1 kc/s); (*d*) characteristic obtained by means of pulse (*c*) : upper curve diode No. 1, lower curve diode No. 2; (*e*) typical current pulses for voltages just below and just above ' pulse breakdown '.

The horizontal voltage scale is similar for (*a*), (*b*) and (*d*). The vertical current scale is extended about three times in case (*d*) compared with (*a*) and (*b*).

when the voltage exceeds a certain critical value.* This value was found by Billig to correspond closely to the a.c. turnover voltage. Our results do not wholly confirm his. This is so not only on account of the indefiniteness of turnover mentioned above, but also because the critical voltage depends upon the duration of the pulse.

§ 2. Experimental

The circuit used for the study of pulse breakdown is shown in figure 2. The square pulse generator supplied pulses of variable duration from 3 to 250 μsec and a repetition frequency range from 0·54 kc/s to 20 kc/s. Pulses shorter than 18 μsec could, however, only be obtained at the higher repetition frequencies. The pulses were applied to the diode via an amplifier used mainly to decrease the output impedance of the generator. The voltage and current were measured respectively at the points Y_1 and Y_2 using a double beam Cossor oscilloscope.

Figure 1 (*e*) shows the superimposed traces of a current pulse just before and just after pulse breakdown. The breakdown starts invariably as a peak at the end of the pulse, independent of its length. The peak grows rapidly in height and breadth with voltage.

* We shall use the term ' pulse breakdown ' to denote this phenomenon in distinction to other forms of breakdown such as voltage turnover.

The first test was to see whether the breakdown is in any sense a cumulative effect of the train of pulses. Selecting an arbitrary pulse width the breakdown voltage was measured as a function of frequency and proved to be constant over a considerable range. For instance, for a pulse width of 40 μsec, the breakdown voltage (and current) remains constant within the accuracy of the experiment from 0·54 kc/s to 10 kc/s. When the pulse length becomes comparable with the interval between pulses, deviations from constancy of breakdown voltage may occur. In all subsequent measurements we therefore ensured that the frequency was as low as possible and such as not to introduce any cumulative effects.

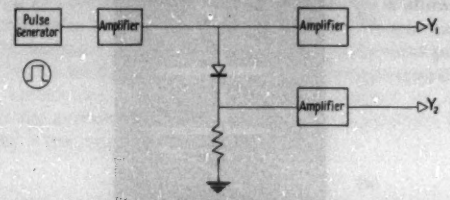


Figure 2. Circuit used for the study of pulse breakdown.

Figure 3 represents the variation of power $P(=IV)$ dissipated at breakdown with pulse width t.† The measurements were taken using diode No. 2 and correspond to ambient temperatures of 20°c and 150°c. The meaning of the broken line will be explained in the following section.

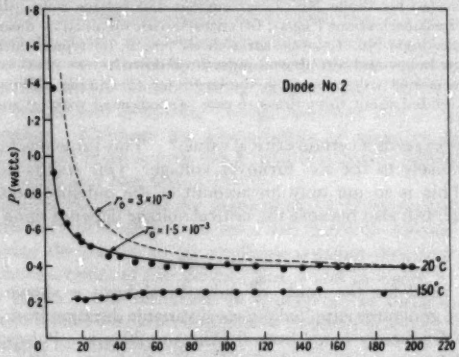


Figure 3. Power necessary to cause breakdown plotted against pulse duration for two ambient temperatures. Curves marked $r_0 = 3 \times 10^{-3}$ and $r_0 = 1{\cdot}5 \times 10^{-3}$ are obtained from theory (see § 3).

Let us consider first the case of room temperature. The result is typical of several diodes studied inasmuch as it shows the steep rise of power necessary to cause breakdown for pulses shorter than, say, 40 μsec. In some cases the

† A plot of breakdown voltage against t would have a similar shape, the initial fall being less steep. The breakdown voltage is 140 volts at $t = 4$ μsec, falling to 90 volts for $t > 60$ μsec.

power continues to decrease somwhat for longer pulses, although at a rate very much slower than the initial fall. At 150°c the power is also essentially constant over most of the range. Measurements with pulses shorter than 18 μsec could not be made because the higher repetition frequencies required did introduce cumulative effects. We do not attribute any real significance to the slight fall of power between 20 and 60 μsec as it may be due to an experimental error connected with the difficulty of measurements at high temperatures discussed at the end of this section.

The magnitude of the current and voltage pulses at breakdown enables us to determine the resistance of the diode. Figure 4 represents the dependence of this resistance on ambient temperature θ_0. The repetition frequency used in these measurements was the lowest possible (0·54 kc/s) and the pulse width 70 μsec. This particular pulse width was chosen for convenience of measurement (see the end of this section) and for reasons which will become apparent in the discussion of these results. Shorter pulses were also tried but did not introduce significantly different results.

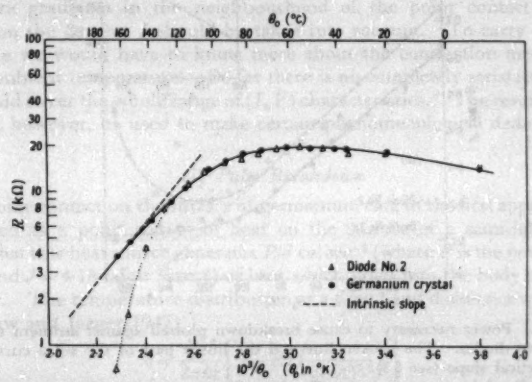


Figure 4. Resistance of diode at breakdown and of germanium crystal plotted against inverse ambient temperature. The resistance scale corresponds to the diode. The resistance of germanium was multiplied by a constant factor to make the two curves coincide at the maximum. Resistivity of Ge used in diode No. 2 $\simeq$ 3 Ω cm.

The diode was suspended in an oil bath heated by an electric heater. The temperature was measured by a mercury thermometer immersed in the oil. A thermocouple attached to the base of the diode was used to check whether any excess heating due to the passage of pulses was present. No difference between the thermometer and thermocouple readings could be observed.

Figure 4 shows the results for diode No. 2. The triangles represent the resistance of the diode and the circles the resistance of the germanium crystal obtained by dismantling the diode and replacing the whisker by a large area soldered electrode. The resistance of the crystal was measured by passing a constant current and measuring the potential difference with a potentiometer. The values of this resistance at different temperatures were multiplied by a constant factor chosen so as to make the maximum (at about 50°c) coincide with

the resistance of the diode at the same temperature. Three diodes were investigated in this manner and gave plots similar to those of figure 4. The point at which the diode resistance curve departs from that of the germanium crystal was in all cases between 100°C and 130°C. The rapid fall of resistance at high temperatures could not be due to heating of the bulk of the germanium specimen above ambient by the passage of pulses, because of the equality of the thermometer and thermocouple readings. It is extremely unlikely that temperature differences up to several tens of centigrade degrees could be sustained by the metal base about 3 mm thick. Besides, such temperature gradients would probably be accompanied by time effects such as creeping currents which were not observed.

Figure 5 (*a*) gives the variation of the power dissipated at breakdown with ambient temperatures for diodes No. 2 and No. 2 a (similar in its (I, V) characteristics to No. 2). A similar plot has been obtained by Benzer (1949). According

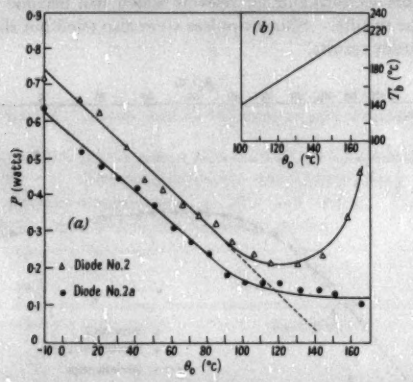


Figure 5 (*a*). Power necessary to cause breakdown plotted against ambient temperature for two diodes. The broken line and the linear part of the solid curve have the theoretical slope (see § 3).

(*b*) Breakdown temperature T_b plotted against ambient temperature obtained from thermal instability theory.

to this author the extrapolation of the linear part of the curve to zero power should give the contact temperature at breakdown (in our case about 147°C). The rise of power at high temperatures also obtained by Benzer is absent for diode No. 2 a. The slope of the linear part of the curve for diode No. 2 was obtained from theoretical arguments (see next section) and was not chosen as the best fit for the experimental points. The curved part of the curve (above 90°C), for which no detailed theoretical explanation is available, follows the experimental points.

Before proceeding with the discussion some remarks about the experimental procedure must be made. Most of the above results depend upon the measurement of current and voltage at the point when pulse breakdown starts to occur. This breakdown, however, is not always as definite as the trace of figure 1 (*e*) would suggest. This is especially true at high temperatures and at small pulse width. The sharp peak at the end of the pulse tends to degenerate into a constant slope extending sometimes throughout the whole length of the pulse.

As the voltage is increased this slope increases and eventually the top of the pulse becomes convex towares the time axis. In doubtful cases we have taken this point as indicating breakdown. It is much easier to make measurements with pulses of about 70 μsec duration than, say, 20 μsec. The relatively slow time base of the Cossor oscilloscope does not make it an ideal instrument for the study of short pulses. The reproducibility of the results is fairly good except in cases when between two measurements at room temperature the diode has been heated to approximately 170°c. The changes which occur may perhaps be attributed to mechanical effects due to the expansion of the diode envelope and consequent movements of the whisker.

§3. Discussion

To obtain a quantitative theory of pulse breakdown a much more thorough knowledge of germanium diodes than we possess at present would be necessary. Ideally such a theory should be able to predict the variation of current I with time, after the application of the constant voltage pulse. Time dependent temperature gradients in the neighbourhood of the point contact and their influence on the current I should be taken into account. To carry out such a programme we would have to know more about the conduction mechanism at constant ambient temperature. So far there is no completely satisfactory theory which would cover the whole range of (I, V) characteristics. The results reported above can, however, be used to make certain phenomenological deductions.

(i) *Pulse Breakdown*

The point contact on the surface of germanium can, in the first approximation, be regarded as a point source of heat on the surface of a semi-infinite solid. Suppose that this heat source generates P/J cal sec^{-1} (where P is the power applied in watts and $J = 4{\cdot}18$ cal w^{-1} sec^{-1}) of heat which flows into the body of the semiconductor. The temperature distribution at a time t and distance r will be given by (Carslaw and Jaeger 1947)

$$\theta = \frac{P/J}{2\pi kr}\left[1-\operatorname{erf}\left(\frac{r}{2(Dt)^{1/2}}\right)\right] \qquad \ldots\ldots(1)$$

where k is the thermal conductivity (0·12 cal sec^{-1} cm^{-1} deg^{-1}), $D = k/\rho\sigma$, ρ the density (5·5 g cm^{-3}) and σ the specific heat (0·074 cal g^{-1} deg^{-1}) of germanium. Since for $r \to 0$ the temperature θ would become infinite we have to assume that the heat is generated not at a point but within a hemisphere of radius r_0. Expression (1) would then be valid for $r \geqslant r_0$.

On application of a power P to the germanium diode the temperature of the periphery of the hemisphere r_0 will rise according to (1). Let us suppose that there exists a certain critical temperature θ_c at which the pulse breakdown will occur. This temperature will be reached at a time t after the application of the pulse. The relation between the power P and the time t will thus be given from (1) by

$$P = \frac{2\pi J\theta_c kr_0}{1-\operatorname{erf}\{r_0/2(Dt)^{1/2}\}} \qquad \ldots\ldots(2)$$

The two unknown constants in this equation are θ_c and r_0. Putting

$$2\pi J\theta_c kr_0 = a(r_0) = 1$$

we can plot P against t for different values of r_0. A family of such curves is given in figure 6 for three different values of r_0. For a *particular* t the value of P can now

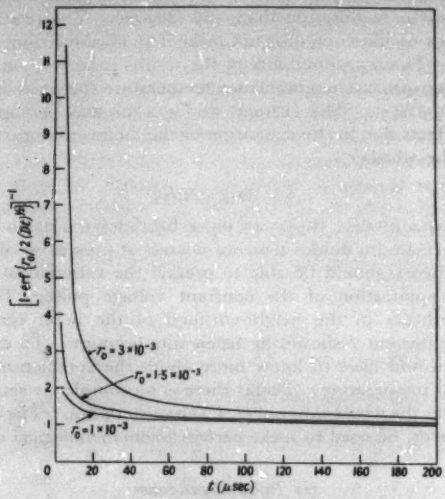


Figure 6. Theoretical curves corresponding to equation (2) for $a(r_0)=1$.

be made to coincide with an experimental point by adjusting the value of $a(r_0)$. Having obtained $a(r_0)$ from a fit at $t=200\,\mu$sec, the two upper curves in figure 3 have been calculated. The broken curve corresponds to $r_0=3\times10^{-3}$ cm, the solid curve to $r_0=1{\cdot}5\times10^{-3}$ cm and fits the experimental points remarkably well. This latter value of r_0 is in good agreement with values obtained by Tillman and Henderson (1953) for the radius of the barrier.

From a knowledge of r_0 and $a(r_0)=0{\cdot}353$, for this curve, we can now calculate the temperature θ_c:

$$\theta_c=\frac{a(r_0)}{2\pi Jkr_0}=75°.$$

It follows that the pulse breakdown should take place at a temperature

$$\theta_b=\theta_c+\theta_0=95°\mathrm{C}$$

where θ_0 is the ambient temperature ($=20°$C). This result is in very close agreement with the value about 90°C obtained by Benzer by a method referred to in the previous section. On application of Benzer's procedure to our results of figure 5 (*a*) we obtain, however, a much higher value, $\theta_b=147°$C.

This discrepancy between our results obtained by the two methods may be due to several reasons. Firstly, our theoretical model is of a rather crude nature since we did not take into account the flow of heat along the whisker and we have assumed tht all the power is dissipated at the contact, thus neglecting the spreading resistance. Secondly, some permanent change may have taken place in the diode

between the measurements of P against t (figure 3) and P against θ_0 (figure 5 (*a*)). This last experiment was carried out by using 70 μsec pulses. Thus the value of P at 20°c at $t=70\,\mu$sec in figure 3 ($\simeq 0{\cdot}4$ watt) should be the same as P for $\theta_0=20$°c in figure 5 (*a*). This latter value is, however, about 1·5 times larger than the previous.

From the theory presented above, and within its approximations, we can deduce the theoretical slope of the linear portion of the curve in figure 5 (*a*) (for diode No. 2). Assuming that for $t=70\,\mu$sec the term $\operatorname{erf}\{r_0/2(Dt)^{1/2}\}$ in (1) is negligible,* θ_c is given by $\theta_c = P/2\pi Jkr_0$, whence

$$\theta_b = \frac{P}{2\pi Jkr_0} + \theta_0. \qquad \ldots\ldots(3)$$

If θ_b is constant for a range of θ_0, a plot of P against θ_0 should have a slope of $-2\pi Jkr_0$. The broken line in figure 5 (*a*) corresponds to this theoretical slope. The agreement with the experimental points is thus quite good, and θ_b seems to remain constant up to $\theta_0 \simeq 90$°c. The whole curve may, however, have been displaced upwards due to the change referred to above.

It could be of interest to repeat the analysis leading to our value of $\theta_b (=95$°c) for a number of different ambient temperatures and determine the variation of θ_b with θ_0. This experiment was not performed, but from the agreement of (3) with experiment it is apparent that θ_b remains approximately constant up to θ_0 of the order of 100°c. This last conclusion is also supported by an analysis of figure 4. Suppose that the resistance R of the diode at the point of breakdown is the sum of two temperature dependent resistances in series: R_c due to the contact (determined by the breakdown temperature θ_b) and R_g due to the bulk of the germanium crystal in thermal equilibrium with its surroundings (thus determined by θ_0):

$$R(\theta_b, \theta_0) = R_c(\theta_b) + R_g(\theta_0). \qquad \ldots\ldots(4)$$

Since $\theta_b =$ constant, the temperature dependence of R_b will be that of $R_g(\theta_0)$. Hence we should expect the germanium sample fitted with large area electrodes to have the same shaped (log resistance, $1/\theta_0$) curve as the diode so long as θ_b remains constant. This is in fact what we observe in figure 4. The departure of the two curves takes place at $\theta_0 \simeq 130$°c, beyond which the resistance of the diode is seen to decrease much more rapidly than the resistance of the crystal. Since at these ambient temperatures we are near the intrinsic conductivity region we can make the reasonable assumption that both R_c and R_g depend exponentially upon the temperature:

$$R(\theta_b, \theta_0) = C_1 \exp(-\phi_1/k\theta_b) + C_2 \exp(-\phi_2/k\theta_0) \qquad \ldots\ldots(4')$$

where ϕ_1 and ϕ_2 are two activation energies. If now θ_b increases with θ_0 the resistance is seen to decrease with temperature faster than if the first term in (4′) were constant.

The similarity between the temperature dependence of the resistance of the diode and germanium below 130°c may have an important meaning. It suggests that the mechanism of conductivity in germanium is the same as in a diode through which breakdown current is flowing. By this we mean that the same proportions of electrons and holes take part in the conduction in both cases. The usual

* This is the second reason for using long pulses in these experiments—see end of § 2.

theories of inverse current in germanium rectifiers take into account one type of carrier only (see Torrey and Whitmer 1948), and the difficulties connected with this picture have been discussed by Bardeen and Brattain (1949).

(ii) *Thermal Instability*

Let us now discuss the phenomenon of thermal instability and see whether it can account for some of our experimental results.

The thermal instability origin of breakdown has been suggested by Billig (1951) in connection with selenium rectifiers and is due to the fact that a slight increase in temperature may cause a larger current to flow and hence a larger power loss which in turn increases the temperature. The condition for stability obtained by Billig may be written in the following form:

$$T-\theta_0 \leqslant P(\partial P/\partial T)_V \qquad \ldots\ldots(5)$$

where T is the operating temperature of the rectifier, θ_0 the ambient temperature, P the power loss and V the applied voltage. To calculate $(\partial P/\partial T)_V$ we must know how $P(=IV)$ depends upon T. In other words we must know the (current, temperature) characteristic of the rectifier for voltages not far from breakdown. Neglecting the flow of carriers against the field (see Simpson and Armstrong 1953) this characteristic will be given by

$$I=I_0 \exp(-\phi/\boldsymbol{k}T), \qquad \ldots\ldots(6)$$

where ϕ is an effective barrier height and I_0 may be a function of V. From (5) and (6) we can obtain the following condition for stability:

$$T-\theta_0 \leqslant \boldsymbol{k}T^2/\phi. \qquad \ldots\ldots(7)$$

Limiting values of T (which we shall call T_b) can be obtained from the quadratic equation by taking the equality sign and are given by*

$$T_b=\phi/2k[1-(1-4\theta_0\boldsymbol{k}/\phi)^{1/2}]. \qquad \ldots\ldots(8)$$

If we want to apply (8) to the case of pulse breakdown we must identify T_b with our barrier temperature θ_b. The important quantity in this expression is the energy ϕ. We do not have any reliable information about the magnitude of the effective barrier height near the breakdown point. It should be noted, however, that when this barrier height is reduced to about $4k\theta_0$ ($\simeq$0·1 ev) thermal instability will not occur, however high the operating temperature θ_b may rise. Some measurements of Simpson and Armstrong indicate that ϕ may be reduced to as low as 0·19 ev. Moreover the suggestion made earlier that at the point of breakdown electrons and holes take part in the conduction process in the same ratio as in bulk germanium means that the barrier is almost completely absent. If we assume $\phi=0{\cdot}15$ ev we obtain from (8) $T_b=78$°C at room temperature and about 270°C at $\theta_0=100$°C. Such a rapid rise would be entirely incompatible with our previous results.

It could perhaps be argued that even when the effective barrier has been reduced to zero the diode would still have the (current, temperature) characteristic given by (6), ϕ denoting now the activation energy for free carriers. This activation energy would be quite small for temperatures below the intrinsic range but would reach 0·38 ev (half the energy gap in Ge) at higher temperatures.

* The minus sign is chosen in front of the square root because on general grounds we would expect T to increase with θ_0.

Figure 5 (*b*) shows a plot of T_b against θ_0 obtained from (8) for this value of ϕ. The rise of T_b is much slower for this higher value of ϕ and could perhaps account for the increase of the breakdown temperature above about 100°C.

§4. Conclusions

Summarizing the results, we can conclude that the pulse breakdown takes place when the periphery of a hemisphere of approximately $1{\cdot}5 \times 10^{-3}$ cm radius situated beneath the point of contact reaches a temperature somewhat below the onset of intrinsic conductivity in germanium. Our estimates from three different arguments are 95°C (equation (2) and figure 3), 130°C (figure 4) and 147°C (figure 5 (*a*)). The discrepancy is probably due to the rather crude model used and to the possibility of mechanical changes in the diode between experiments. The breakdown temperature remains essentially constant from $\theta_0 = -10$°C to $\theta_0 \simeq 100$°C, whereupon it starts to increase. These results are in reasonable agreement with the measurements of Tipple and Henisch (1953), who determined the contact temperature by thermoelectric methods.

Thermal instability is certainly not the cause of breakdown for $\theta_0 \lesssim 100$°C. For $\theta_0 \gtrsim 100$°C the question remains open and would require more detailed investigation.

§5. Acknowledgments

The authors are indebted to Dr. C. S. Bull for reading and discussing the manuscript, to Messrs. E. Grodzinsky, Hall and Styles for designing and making most of the electronic equipment, and to Mr. B. Wilkinson, Managing Director of Electronic Tubes Limited, for permission to publish this paper.

References

Bardeen, J., and Brattain, W. H., 1949, *The Bell Syst. Tech. J.*, **28**, 239.
Benzer, S., 1949, *J. Appl. Phys.*, **20**, 804.
Billig, E., 1951, *Proc. Roy. Soc.* A, **207**, 156; 1952, *Phys. Rev.*, **87**, 1060.
Carslaw, H. S., and Jaeger, J. C., 1947, *Conduction of Heat in Solids* (Oxford: Clarendon Press).
Hunter, L. P., 1951, *Phys. Rev.*, **81**, 151.
Simpson, J. H., and Armstrong, H. L., 1953, *J. Appl. Phys.*, **24**, 25.
Tillman, J. R., and Henderson, J. C., 1953, *Phil. Mag.*, **44**, 677.
Tipple, P. M., and Henisch, H. K., 1953, *Proc. Phys. Soc.* B, **66**, 826.
Torrey, H. C., and Whitmer, C. A., 1948, *Crystal Rectifiers*, Part I (New York: McGraw-Hill).

Determination of Relaxation Spectra from Damping Measurements

By F. C. ROESLER and J. R. A. PEARSON
Butterwick Research Laboratories, Imperial Chemical Industries Limited, Welwyn, Herts.

Communicated by H. Kolsky; MS. received 9th December 1953

Abstract. A method is shown for the determination of the relaxation spectrum from the data furnished by dynamical experiments, and in particular from the damping (imaginary part of the complex modulus) as a function of frequency. The circular functions are eigenfunctions of the relevant integral equation, which can therefore be solved numerically by expansion. The spectrum is obtained by first making a Fourier analysis, multiplying every coefficient by the associated eigenvalue, and then synthesizing the new Fourier series. An estimate is given of the resolving power of the method, which is limited only by the error in the experimental data. As an illustration, the spectrum of polyisobutylene in the frequency domain corresponding to the rubber–glass transition is found from published data.

§1. Introduction

The mechanical properties of materials like high polymers can be investigated experimentally by subjecting specimens to periodic deformations. Since high frequencies can be used, the method is especially suitable for gaining information about processes which take place rapidly.

For stresses and strains which are not too large the visco-elastic properties exhibited in such an experiment can be described by a complex Young's modulus E (or shear modulus G or bulk modulus K depending on the type of deformation) which is a function of the pulsatance Ω:

$$E(\Omega)=E'(\Omega)+iE''(\Omega). \qquad \ldots\ldots(1)$$

This implies that for the domain of stresses and strains in question there exists a linear relation

$$\sigma=\epsilon E \qquad \ldots\ldots(1\,a)$$

where both the stress and the strain are harmonic functions of time :

$$\sigma=\sigma_0 \exp(i\Omega t), \quad \epsilon=\epsilon_0 \exp(i\Omega t), \quad (\sigma_0 \text{ and/or } \epsilon_0 \text{ complex}). \qquad \ldots\ldots(1\,b,c)$$

The complex modulus can be related to the relaxation time or to the spectrum of relaxation times of the material. For a material with a single relaxation time T we have

$$E' \propto \frac{\Omega^2T^2}{1+\Omega^2T^2}, \qquad E'' \propto \frac{\Omega T}{1+\Omega^2T^2}. \qquad \ldots\ldots(2\,a,b)$$

In the case of a spectrum of relaxation times $\phi(T)$, the analogous relations are

$$E'(\Omega)=\int_0^\infty \phi(T)\frac{\Omega^2T^2}{1+\Omega^2T^2}\,dT, \qquad E''(\Omega)=\int_0^\infty \phi(T)\frac{\Omega T}{1+\Omega^2T^2}\,dT. \qquad \ldots\ldots(3\,a,b)$$

If E is measured in suitable units, ϕ is a normalized distribution function

$$\int_0^\infty \phi(T)\,dT=1. \qquad \ldots\ldots(4)$$

In what follows we shall not enter into a discussion of the interpretation of the spectrum. It is by no means straightforward. The analysis of the visco-elastic behaviour into a distribution of 'Maxwell elements', which the spectrum implies, is a formal one; these 'elements' need not and, in general, will not, be identifiable as physical units or structures in the material. For perhaps the most successful attempt so far to elucidate this obscure field we refer the reader to the pioneer investigations by Kuhn, Künzle and Preissmann (1947), where the derivation of our formulae (2) and (3), as well as further references, can also be found. We shall be concerned here merely with the mathematical problem of finding the spectrum from a given damping function $E''(\Omega)$. Various authors have evolved procedures to determine the spectrum from observable quantities by 'approximation' methods which replace the basic equations by others: for instance, the kernels in the integral equations (3 *a*, *b*) are replaced by certain step functions and delta functions. For an able review of this work see Leaderman (1953). However, there is no need thus to depart from the basic equations. The scheme described below for finding the spectrum is, we believe, more systematic and numerically more powerful. Admittedly, however, our method requires more computation.

As would be expected, the limit of the actual resolving power in our (in principle exact) method is given by the accuracy and the range of the measurements made in determining the damping function. The latter, however, is very insensitive to small local variations in the amplitude of the spectrum function. The integration (3 *b*) which leads from the spectrum to the damping function is responsible for this: it averages, and thereby irons out all details of small extension along the log Ω-scale. Therefore, as we shall see, the experimental accuracy in the damping measurements required to make a given detail of the spectrum recognizable increases exponentially with increasing fineness of the detail.

The method described below should be adaptable to the numerical solution of other integral equations if they are of the convolution type (cf. below) or if they become so after a suitable transformation of the variables. It would appear that many problems in the theory of visco-elasticity can thus be attacked.

§2. Solution of the Integral Equation

We change the variables in (3 *b*) from T and Ω to $\ln(T/T_0)$ and $-\ln(\Omega T_0)$. An alternative, and with respect to all that follows an equivalent, change is to $-\ln(T\Omega_0)$ and $\ln(\Omega/\Omega_0)$. *We shall use the letters τ and ω to denote (either set of) the new, logarithmic, variables.* The equation (3 *b*) now reads

$$E''(\omega) = \tfrac{1}{2}\int_{-\infty}^{+\infty} \phi(\tau)\,\mathrm{sech}\,(\omega-\tau)\,d\tau \qquad \ldots\ldots(5)$$

where the letter ϕ now stands for a distribution function referring to the variable τ. The integral equation (5) is one 'of the first kind' and its kernel depends upon the difference $\omega-\tau$ only. For kernels of this ('convolution') type the eigenfunctions and the associated eigenvalues are accessible by a known, general method (cf. Titchmarsh 1937). It is shown in the literature that by subjecting the general form of the convolution type of equation, with ψ an eigenfunction,

$$\psi(s) = \lambda \int K(s-t)\psi(t)\,dt \qquad \ldots\ldots(6)$$

to Fourier transformation, one arrives at an equation

$$1-\lambda\int K(\xi)\exp(i\nu\xi)\,d\xi = 0 \qquad \ldots\ldots(7)$$

from which the eigenvalues λ may be found. It is also shown that to the eigenvalues $\lambda(\nu)$ corresponding to all simple zeros of (7) are associated eigenfunctions of the type $\exp(i\mu s)$ such that $\mu=\nu$. For the kernel $\frac{1}{2}\operatorname{sech}\xi$ of (5) equation (7) becomes

$$\lambda-\frac{2}{\pi}\cosh(\nu\pi/2)=0. \qquad \ldots\ldots(8)$$

(To find the Fourier transform of sech ξ one may remember that sech $[x\sqrt{(\pi/2)}]$ is a self-reciprocal function with regard to Fourier transformation.) If (7) has multiple zeros then the eigenvalues $\lambda(\nu)$ corresponding to these are associated with certain special eigenfunctions (cf. Titchmarsh 1937). Our form of (7) (i.e. (8)) possesses, in addition to an infinity of simple zeros, one double zero, namely at $\nu=0$. The corresponding eigenvalue is $2/\pi$, and it is associated with eigenfunctions of the form $\psi=A\omega+B$.

The integral equation (5) is 'singular', the domain of integration being infinite. It is a consequence of the singular character of this equation that the eigenfunctions form a continuous manifold and are not all mutually orthogonal. For a singular equation one cannot, in general, make definite assertions about the possibility of solution by expansion. Formally, we may in any linear integral equation of the first kind

$$f(s)=\int K(s,t)g(t)\,dt \qquad \ldots\ldots(9)$$

expand $f(s)$ and $g(t)$ in eigenfunctions ψ_m:

$$f(s)=a_1\psi_1+a_2\psi_2+\ldots+a_m\psi_m+\ldots \qquad \ldots\ldots(10a)$$

$$g(t)=b_1\psi_1+b_2\psi_2+\ldots+b_m\psi_m+\ldots \qquad \ldots\ldots(10b)$$

Using the definition of an eigenfunction the equation becomes

$$\sum a_m\psi_m(s)=\sum(b_m/\lambda_m)\psi_m(s) \qquad \ldots\ldots(11)$$

If the ψ_m are mutually orthogonal, the only solutions of (11) are

$$b_m=\lambda_m a_m. \qquad \ldots\ldots(12)$$

Even if the ψ_m are not orthogonal, (12) gives a set of solutions. There remains then the question of the significance of the developments (10 a, b). Convergence, or at least approximation in the mean, must be proved. Since we know that in our case the circular functions $\exp(i\nu\omega)$ with arbitrary ν are eigenfunctions we can fall back on the general theory of the Fourier integral and the Fourier series. From this we know that suitably chosen sets of circular functions $\exp(i\nu\omega)$ form complete systems of functions with respect to arbitrary chosen domains of ω, and that any quadratically integrable, continuous and sectionally smooth function can be represented by a uniformly and absolutely convergent Fourier series or integral. Thus any empirically given $E''(\omega)$ can be expressed as a sum or integral of eigenfunctions of (5): the development (10 a) is, in our case, always possible. This does not imply that (10 b) will also converge, but if a solution to the integral equation exists (cf. §4 below) it must be given by (10 b).

§3. Procedure for Numerical Analysis; Estimation of the Resolving Power

If $E''(\omega)$ is measured over a sufficiently wide range of ω, i.e. one which includes the whole relevant relaxation spectrum of the material, E'' will tend to zero at the limits of this domain. In such cases the spectrum is completely determined by the measurements (within the experimental accuracy). We shall call this case 1.

If $E''(\omega)$ has only been measured within a limited range, such that E'' does not tend to zero at its limits, but has finite, possibly large, values there, then, quite apart from the question of experimental accuracy, the measurements do not, in principle, determine the spectrum. However, if the range covered by experiments is fairly extensive it may be possible to derive from the measured $E''(\omega)$ a spectrum, and this is likely to approximate to the true spectrum at least in the central region of the experimental range. Our method can be used for this purpose, but it cannot, of course, supply the missing information with regard to the damping in the unexplored domains. We shall call this situation, with data of insufficient coverage, case 2.

If case 2 arises it can, for the purpose of the subsequent analysis, be formally reduced to case 1 by subtracting a suitable $A\omega + B$ type eigenfunction from $E''(\omega)$, thus making the remainder of $E''(\omega)$ vanish at the limits of the experimental range. Various possibilities exist for trying to improve on this by guesswork. For instance, before the said subtraction we may extrapolate the experimental $E''(\omega)$ for a few units of ω beyond the experimental limits, and then work with this extended domain. It may be noted that in view of the periodicity assumption inherent in the subsequent treatment the method of subtracting a linear function also implies an extrapolation. In what follows we assume that we are faced with case 1 proper or that the assumptions necessary for the reduction of a case 2 have already been supplied. Therefore we now consider case 1, which is mathematically definite.

The damping function $E''(\omega)$ can then, over the relevant range, be represented by a series of sine terms :

$$E''(\omega) = a_1 \sin x + a_2 \sin 2x + \ldots a_k \sin kx + \ldots a_n \sin nx \quad \ldots\ldots (13)$$

with a_n the last coefficient given with any accuracy by the experimental data for $E''(\omega)$. To make this representation possible the new variable x should be a linear function of ω such that $x<0$ and $x>\pi$ correspond to values of ω outside the domain of ω over which $E''(\omega)$ is significantly different from zero. (The test of significance is here by comparison with the absolute error of E'' within the domain, not with the relative error at the proposed end points.) $E''(\omega)$ may then be assumed to be periodic $[E''(x+2\pi) = E''(x)]$ and odd $[E''(x) = -E''(-x)]$ without thereby exceeding in the domain $0<x<\pi$ the allowance given by the experimental error. The first property makes the analysis into a Fourier series possible, the second makes all cosine terms drop out. This choice of the basic domain for the sine analysis must be made in the light of the circumstances of any desired application; in principle it is clear that the errors brought in by the periodicity assumption can be made arbitrarily small in the treatment of any case 1 proper. When making the choice it is desirable to take into account the probable rapidity of the convergence of the series (13), which depends on the length of the domain (see below). In the sine analysis all components well down into noise level should be determined. (For large k the values of the components as found numerically will not tend to zero, but to a random spectrum of finite amplitude. This is the 'noise' due to the various errors.)

Supposing the a_k to have been found, then from the preceding section the spectrum is, within the domain $0<x<\pi$, given by

$$\phi(x) = b_1 \sin x + b_2 \sin 2x + \ldots b_n \sin nx \quad \ldots\ldots (14)$$

where $b_k = a_k (2/\pi) \cosh(\nu\pi/2)$. The ν in this formula comes from $\exp(i\nu\omega)$ or

sin $\nu\omega$. If, say, a length L of the ω-domain is mapped on to $0<x<\pi$, we have $\nu = k\pi/L$ and

$$b_k = a_k \frac{2}{\pi} \cosh(k\pi^2/2L). \qquad \ldots\ldots(15)$$

We now consider what error of the a_k is admissible if a specified resolution is to be obtained. For example, we wish to know the accuracy necessary to bring out a detail of extension ϵ along the ω-scale and of amplitude η relative to the mean amplitude of the spectrum. The values of ν in the Fourier series must then run up to $\nu = \pi/\epsilon$ (at least) and the factor which relates the corresponding a and b is $(2/\pi)\cosh(\pi^2/2\epsilon)$. This b will be smaller, by a factor approximately equal to η, than the largest b in the series. Thus, in order that it should be possible to determine the critical a corresponding to $\nu = \pi/\epsilon$, the error in the a_k, expressed as a percentage of the largest a, must satisfy

$$\Delta a \lesssim 50\,\pi\eta\,\mathrm{sech}(\pi^2/2\epsilon). \qquad \ldots\ldots(16)$$

This gives

$$\Delta a \lesssim 5\% \quad \text{for } \epsilon = 2, \quad \eta = 0{\cdot}2$$
$$\Delta a \lesssim 0{\cdot}5\% \quad \text{for } \epsilon = 1, \quad \eta = 0{\cdot}2$$
$$\Delta a \lesssim 0{\cdot}003\% \quad \text{for } \epsilon = 0{\cdot}5, \eta = 0{\cdot}2$$

These ϵ-values must be divided by ln 10 to find their equivalents in decades. Thus $\epsilon = 1$, above, stands for 0·43 decade.

If we think that the resolution characterized by a given value of ϵ can and should be obtained in a given application, then the number of coefficients in the series (13) which must be determined is given by $n = L/\epsilon$. It seems to us unlikely that in any application the domain of frequencies to be covered will exceed 16 decades, and from the figures above, with present experimental techniques, a resolution of approximately 0·5 decade is about the best one can hope for. For these reasons we have found it convenient to standardize the procedures for the sine analysis such that every analysis is made for 31 coefficients, using 32 ordinates, of which one is zero by definition.

§4. Properties of the Damping Function necessary to make the Equation Soluble

The factor $(2/\pi)\cosh(k\pi^2/2L)$, which by (15) relates the two series (13) and (14), increases rather rapidly with increasing k. In order that the series for the spectrum (14) should nevertheless be convergent

$$\text{(in the sense that } \lim_{k\to\infty} \sum_k b_k^2 = N, \text{ a finite norm)},$$

the a_k in the series for the damping function (13) must decrease even more rapidly. This condition will not, in general, be fulfilled by an arbitrary function, even if the term arbitrary be qualified to exclude obviously pathological functions. If a function which does not satisfy this condition is taken for the left-hand side in (5), then the integral equation will not possess a convergent solution. This situation, where an integral equation of the first kind turns out not to possess solutions for arbitrary $f(s)$ is by no means due to a peculiarity of (5); it is the rule rather than the exception (cf. Courant–Hilbert 1931, Vol. I, chapter III, para. 10, section 11). We here recall a theorem which gives upper bounds for the values of Fourier coefficients. It asserts (cf. Courant–Hilbert, 1931, Vol. I, chapter II, para. 5,

section 3) that if the function to be developed and its derivatives up to order $h-1$ are continuous, and if the derivative of order h is sectionally continuous, the coefficients obey

$$|a_\nu| < C/\nu^h \qquad \ldots\ldots(17)$$

with C a constant. By contrast, for (14) to converge we require

$$|a_k| < C_1/\cosh(kC_2). \qquad \ldots\ldots(18)$$

Since cosh x increases in the limit of large x more quickly than any finite power of x, any function which makes the equation soluble and all its derivatives must be continuous, up to infinitely high order.

To these conditions on $E''(\omega)$ which stem from the structure of the integral equation must be added others due to the physical restrictions for the relaxation spectrum. This must not be negative anywhere and it must be normalized or normalizable. The kernel $\frac{1}{2}\operatorname{sech}(\omega-\tau)$ in (5) is a continuous function and infinitely many times differentiable. If we differentiate (5) n times with regard to ω we have

$$\frac{d^n E''(\omega)}{d\omega^n} = \tfrac{1}{2}\int_{-\infty}^{+\infty} \phi(\tau)\, d\tau \frac{d^n \operatorname{sech}(\omega-\tau)}{d\omega^n}. \qquad \ldots\ldots(19)$$

Now the derivatives of sech x are functions of bounded variation and assume some maximum value; we shall call this maximum value of the nth derivative M_n. Since $\phi(\tau)>0$ and $\int\phi(\tau)d\tau = 1$ (say), we then have

$$\frac{d^n E''(\omega)}{d\omega^n} \leqslant \tfrac{1}{2} M_n. \qquad \ldots\ldots(20)$$

This proves that all the derivatives of a possible damping function (i.e. a function generated by the kernel $\frac{1}{2}\operatorname{sech}(\omega-\tau)$ from a physically admissible spectrum) are not only finite but possess certain universal upper bounds. A simple example of a function which cannot represent a true damping function because it violates (20) and for which, therefore, no solution positive throughout to the integral equation (5) exists, is given by $F(\omega) = \exp\{-(\omega-\tau)^p\}$ with p an even integer larger than or equal to 2. This function possesses a maximum at $\omega=\tau$ which is sharper than that of $\frac{1}{2}\operatorname{sech}(\omega-\tau)$, the damping function corresponding to a single line spectrum. Intuitively it is clear that no spectrum satisfying the requirement $\phi(\tau)>0$ and corresponding to the fictitious damping function $F(\omega)$ can be found, since the width of the spectrum cannot decrease to less than zero.

A certain amount of experimental error of the random type is inherent in any empirical damping function. Systematic errors may also be present. Neither of these 'error components' of $E''(\omega)$ will, in general, conform to conditions (18), (20), and this may entail that the solution of the equation diverges or is not positive throughout. There is thus a definite risk that the high-frequency part of the error component, unless detected and eliminated, will become amplified on going over to the spectrum and falsify it.* The development method proposed here makes it easy to overcome this difficulty, since all that is required is an estimate of how many

* Procedures to find approximations to the spectrum by using derivatives of the empirical curves, as proposed by many authors, must involve steps equivalent, more or less, to (15), and thus give greater prominence, in the spectrum, to the high-frequency component (the higher derivatives) of $E''(\omega)$. At the same time these procedures afford no check on the role of the error component in the transition from damping function to spectrum. They are somewhat objectionable for this reason and could only be rigorously justified (as approximations) if special smoothing procedures, based on the theory of the relevant integral equation, were adopted in conjunction with them.

of the Fourier coefficients a_k are significant. From the physical hypothesis that the spectrum exists (i.e. (14) converges) one expects the absolute values of b_k to decrease, broadly speaking, with increasing k. By inspection of the sequence of b_k as computed from a_k, using (15), one will find that up to a certain point this does happen. On reaching the point where b_k again begins to increase systematically one should generally find that the corresponding a_k values are down to noise level. It is then permissible to cut off the remainder of the series (14), and even to smooth the last few of the remaining terms, provided that in doing this one does not exceed the allowance given by the error in the corresponding a_k.

§5. Example: The Rubber–Glass Transition of Polyisobutylene

The data due to a cooperative research project which have been published by Marvin (1953) show that for polyisobutylene, which is a typical rubber-like high polymer, G'', the imaginary part of the dynamical shear modulus, possesses a pronounced maximum at $\log_{10}T=-8$. In the domains $\log_{10}T<-11{\cdot}2$ and $\log_{10}T>-3{\cdot}2$ the damping is, within the experimental accuracy, vanishingly small compared with its values in the region of the maximum. All this refers to a temperature of 25°C.

Table 1. Values of damping function $G''(\log_{10}T)$ of polyisobutylene at 25°C as used for Fourier (sine) analysis. The data are a selection from the material published by Marvin (1953).

Ordinate No.	$-\log_{10}T$	$\log_{10}G''$ (Marvin)	G''†	x	Ordinate No.	$-\log_{10}T$	$\log_{10}G''$ (Marvin)	G''†	x
0	1·6		0	0	16	8·0	9·650	447	$\pi/2$
1	2·0		0		17	8·4	9·520	331	
2	2·4		0		18	8·8	9·276	189	
3	2·8	6·615	0		19	9·2	8·954	90	
4	3·2	6·880	1		20	9·6	8·557	36	
5	3·6	7·150	1		21	10·0	8·160	15	
6	4·0	7·425	3		22	10·4	7·760	6	
7	4·4	7·690	5		23	10·8	7·360	2	
8	4·8	7·980	10	$\pi/4$	24	11·2	6·960	1	$3\pi/4$
9	5·2	8·245	18		25	11·6	6·560	0	
10	5·6	8·530	34		26	12·0	6·160	0	
11	6·0	8·810	65		27	12·4		0	
12	6·4	9·070	118		28	12·8		0	
13	6·8	9·300	200		29	13·2		0	
14	7·2	9·490	309		30	13·6		0	
15	7·6	9·625	422		31	14·0		0	
					32	14·4		0	π

† G'' (units of 10^7 dyn cm^{-2}).

The quoted data show that the problem of determining the spectrum in the region $-3{\cdot}2>\log_{10}T>-11{\cdot}2$ is a case 1 in the terminology of § 3 and the application of our method is here quite straightforward. Marvin has tabulated the data so that values for the damping are given at points with 0·4 decade interval. By taking the domain for the sine analysis to equal 32 such intervals and to run from $\log_{10}T=-1{\cdot}6$ to $\log_{10}T=-14{\cdot}4$ we bracket the region within which G'' is significantly different from zero. It is convenient to use the values of G'' as measured in 10^7 dyn cm^{-2} units as ordinates for the analysis. Table 1 summarizes the data used. Table 2 shows the Fourier (sine) coefficients of the damping function, the

associated eigenvalues and the coefficients of the spectrum which are the products. The actual analysis was done by using forms based on the method of Danielson and Lanczos (1942) whereby the analysis for 32 coefficients is split into four, each involving eight coefficients. This is perhaps the most rapid method known except for those requiring special equipment, and, using it, a sine analysis for 32 coefficients need take only one hour, approximately. It is seen that the Fourier

Table 2. Sine coefficients of damping function (a_k) and of relaxation spectrum (b_k) for polyisobutylene, as evaluated from the ordinates and with the choice of domain indicated in table 1. The factor $(2/\pi)\cosh(\nu\pi/2)$ which converts the a_k into the b_k is the eigenvalue associated with each eigenfunction $\exp(i\nu\omega)$ of equation (5) ($\nu\pi/2 = k\pi^2/2L\ln 10$) with $L = 12{\cdot}8$ (decades).

k	a_k	$\frac{2}{\pi}\cosh\frac{\nu\pi}{2}$	b_k	Rounded off b_k	
				k odd	k even
1	139·81	0·645	90·31	90	
2	17·75	0·672	11·94		12
3	−112·31	0·718	−80·70	−81	
4	−23·13	0·784	−18·14		−18
5	77·00	0·872	67·15	67	
6	18·94	0·982	18·61		18
7	−49·19	1·122	−55·21	−55	
8	−12·06	1·288	−15·54		−15
9	29·75	1·497	44·57	44	
10	7·00	1·746	12·22		12
11	−16·25	2·045	−33·26	−32	
12	−3·75	2·395	−8·98		−8
13	9·63	2·816	27·12	25	
14	1·75	3·316	5·80		5
15	−5·88	3·904	−22·94	−20	
16	−1·25	4·603	−5·75		−4
17	3·38	5·44	18·36	15	
18	1·00	6·41	6·41		2
19	−2·25	7·56	−17·03	−12	
20	−1·13	8·94	−10·06		
21	1·50	10·55	15·83	8	
22	0·50	12·41	6·21		
23	0·37	14·74	5·46		
24	−0·31	17·39	−5·40		
25	−0·31	20·50	−6·63		
26	−0·06				
27	−0·50				
28	−0·25				
29	0·06				
30	0·00				
31	− 0·19				

series of the damping function converges well and that it can be terminated at the 21st coefficient. Recalculation of the ordinates from this truncated series resulted in practically complete agreement (better than within 1% of maximum amplitude) with the original ordinates. The terms in the series beyond the 21st are noise and give an indication of the random error in the data and of the rounding-off errors accumulated during the analysis. The amplitude of the noise is less than 1% of the first coefficients. This small value is presumably due to Marvin's

smoothing of the raw data and does not necessarily imply an absence of larger, systematic errors. In the series which defines the spectrum the b_k reach their minimum values at $k=18$ (even k terms) and $k=23$ (odd k terms). This is near enough to the k for which a_k reaches noise level.

The most reasonable approximation to the spectrum will be the smoothest function compatible with the deduced coefficients. One wants the spectrum to show all details definitely established by the analysis, but would rather lose some doubtful detail than keep spurious oscillations due to errors. Accordingly the absolute values $|b_k|$ should be diminished towards the end of the series by amounts corresponding to the probable error in the a_k, i.e. by amounts given by the noise amplitude of the a_k multiplied by the respective eigenvalue. The smoothed and rounded off b_k adopted and used to find the spectrum as a function of $\log_{10}T$ are shown in the last column of table 2. From its Fourier representation the spectrum can be found by straightforward summation of the components. If one is content with 32 ordinates, this synthesis of the spectrum can also be done by a procedure formally identical with a sine analysis for 32 coefficients, e.g. using again the method of Danielson and Lanczos. (The inversion of a sine transform is again a sine transform.) We have used this method and also computed the values of ϕ for some intermediate points in the central region of the domain by direct summation. The results are given in table 3. For comparison the same table shows the values of the damping multiplied by $2/\pi$, which function may be considered as a first 'approximation' to the spectrum (cf. for example Leaderman 1953). The interpolated values of the damping were found from its Fourier representation.

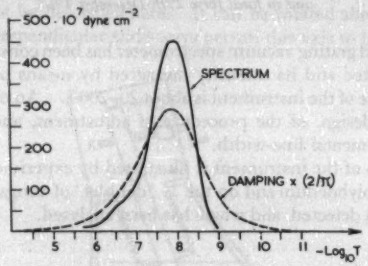
Table 3. Relaxation spectrum ϕ and scaled-down damping function $(2/\pi)G''$ for polyisobutylene. The partly negative 'wings' of the spectrum must be regarded as spurious; their amplitudes are well within experimental error.

$-\log_{10}T$	ϕ (units of 10^7 dyn cm^{-2})	$(2/\pi)G''$	$-\log_{10}T$	ϕ (units of 10^7 dyn cm^{-2})	$(2/\pi)G''$
3·6	−8	1	7·6	370	272
3·8			7·8	443	289
4·0	1	2	8·0	449	285
4·2			8·2	377	257
4·4	6	3	8·4	254	214
4·6			8·6	129	164
4·8	2	6	8·8	41	119
5·0			9·0	5	83
5·2	10	11	9·2	3	55
5·4			9·4		36
5·6	10	22	9·6	8	23
5·8		28	9·8		
6·0	13	40	10·0	−12	10
6·2		57	10·2		
6·4	52	74	10·4	− 1	4
6·6	68	97	10·6		
6·8	85	124	10·8	7	1
7·0	115	159	11·0		
7·2	175	197	11·2	− 4	1
7·4	267	237			

One check on the spectrum is afforded by the normalization condition, which in our case reads

$$\int_{-\infty}^{+\infty} \phi(\log_{10} T)\, d(\log_{10} T) \simeq \frac{2}{\pi} \int_{-\infty}^{+\infty} G''(\log_{10} T)\, d(\log_{10} T). \quad \ldots\ldots (21)$$

This is an exact relation (it has nothing to do with considering $(2/\pi)G''$ as an 'approximation' to ϕ), to be derived from the basic integral equation (5) by integration. The validity of (21) was tested by summing of 32 ordinates for each function and agreement within 1% was found.



The data of table 3 are plotted in the figure. It was felt that such a plot might be desirable, since it may enable an experimenter to see at a glance whether it is worth while embarking on the computations which the application of our method necessitates.

Acknowledgments

We wish to record our indebtedness to Dr. H. Kolsky and Mr. J. J. Benbow for directing our attention to the problem in hand, for advice, for criticism, and for much information on previous work.

References

Courant, R., and Hilbert, D., 1931, *Methods of Mathematical Physics*, Vol. I (Berlin : Springer).

Danielson, G. C., and Lanczos, C., 1942, *J. Franklin Inst.*, **233**, 365.

Kuhn, W., Künzle, O., and Preissmann, A., 1947, *Helv. Chim. Acta*, **30**, 307, 464.

Leaderman, H., 1953, *National Bureau of Standards Report* No. 2672.

Marvin, R. S., 1953, *National Bureau of Standards Report* No. 2305 and addendum.

Titchmarsh, E. C., 1937, *Introduction to the Theory of Fourier Integrals* (Oxford : Clarendon Press).

A Geiger Counter Vacuum Spectrometer and its Use for the Study of Soft X-Ray Lines

By J. L. ROGERS* and the late F. C. CHALKLIN
Physics Department, Canterbury University College, Christchurch, New Zealand

Communicated by E. N. da C. Andrade; MS. received 4th June 1953, and in final form 29th December 1953

Abstract. A ruled grating vacuum spectrometer has been constructed in which the radiation is detected and its intensity measured by means of a Geiger counter. The working range of the instrument is about 20–200 Å. An outline is given of the principles of its design. of the procedure of adjustment, and of the method of finding the instrumental line-width.

The operation of the instrument is illustrated by experiments on the shape of the M_ζ line of molybdenum and on the N 'doublet' of tungsten, in which a new structure has been detected, and which has been analysed.

§ 1. Introduction

The determination of the width and shape of soft x-ray lines and bands, and the complete study of the absorption spectra, require intensity as well as wavelength measurements. Wavelengths may be measured with precision by photographic instruments, which are also excellent for the detection of fine structures. But there is always difficulty in establishing the precise relation between the intensity and the photographic blackening. In a preliminary note (1951) we announced the construction of a concave grating vacuum spectrometer which employs a Geiger counter to measure the intensity, and the present paper describes the instrument and gives an account of the first results obtained with it. The instrument is useful for wavelengths down to about 20 Å, but an upper limit of about 200 Å is at present imposed by the absorption of the x-rays by the window of the counter. Piore, Kingston, Gyorgy and Harvey (1951, 1952) have recently built a spectrometer of different mechanical design and employing a Be–Cu photo-electron multiplier. It has been used with great success at wavelengths of several hundred ångström units. The two spectrometers appear to be complementary. The Piore instrument is excellent in the longer wavelength portion of the soft x-ray region, and the present instrument is possibly preferable below 100 Å where narrow slits must be used and where the mechanical requirements for maintaining a good focus are particularly stringent.

§ 2. The Mechanical Design of the Spectrometer

The instrument employs a glass grating which has a radius of curvature of 99·43 cm and was ruled by the Siegbahn engine with 5760 lines per cm over a width of 2 cm.

* Now at the National Research Council, Ottawa, Canada.

A simplified horizontal section of the spectrometer is shown in figure 1. The instrument fulfils the following requirements: (*a*) the angle of incidence on the grating is large (87½°) in order to give good reflected intensity and reasonable dispersion, (*b*) the diffracted X-rays are selected by a narrow second slit S_2, and this slit is movable along the (horizontal) Rowland circle of focus, (*c*) the position on the circle of this slit is determinable with precision, (*d*) the Geiger counter is so mounted that it *always* detects the radiation passing through this slit, (*e*) the optical system is enclosed in an evacuated chamber.

The grating holder resembles that of Magnusson (1938), and the details of its construction will not be described here. It can be rotated about a vertical axis, and two mutually perpendicular slide-ways permit this axis to be established in a

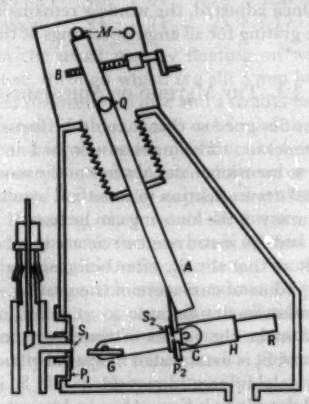


Figure 1. Horizontal section of spectrometer (simplified). G=grating. S_1 and S_2='first' and 'second' slits. P_1 and P_2='first' and 'second' slit plates. Q=bearing of arm A. A=arm carrying second slit. C=counter. H=counter carriage. R=rail for counter carriage. B=screw controlling rotation of arm A.

desired position. Mechanisms are provided to bring the surface of the grating up to this axis and to rotate the grating about horizontal axes parallel and perpendicular to its surface.

The first and second slits are of unusual design. In each slit the blades are held against a flat plate by separate holders, and each holder can be rotated in the plane of the plate so that the blades may be made parallel to one another and so that the slit may be rotated. The positions as well as the widths of the slits may be altered by independent movement of the blades within the holders. In practice the slit-widths were both 10^{-3} cm. The first slit plate P_1 is mounted in the wall of the vacuum box, and the plate and its attached slit may be removed to facilitate the adjustment of the spectrometer. Hand-scraped surfaces and two taper pins ensure its accurate replacement.

The second slit plate P_2 is mounted on the end of the arm A and the distance of the slit from the bearing Q of the arm can be adjusted. When the arm is rotated about its bearing the second slit moves along an arc of a circle which will be called

the 'instrument circle'. The movement is brought about by rotation of the screw B outside the vacuum chamber, flexible stainless steel bellows providing the vacuum seal. The angular setting of the arm is determined independently by the reading of a micrometer which, together with a distance block device, measures the distance M between a point on the arm and a fixed point at the same distance from the bearing Q.

The counter C is provided with a slit-shaped window which is only slightly wider than the beam of x-rays which has passed from the grating and through the second slit. It is necessary that these x-rays should always enter the counter. Accordingly the counter is mounted on a carriage H which may move along the rail R. The rail is pivoted on the vertical axis of the grating holder and the carriage is pivoted on the slit plate P_2 immediately below the slit. The two pivots lie on the axis of the rail. Once adjusted, the window remains collinear with the second slit and the pole of the grating for all angular settings of the arm.

§ 3. The Method of Adjustment

The instrument was designed so that it could be focused without making trial-and-error x-ray experiments. The method adopted is somewhat complicated, and it may be desirable to mention in detail a few of the more special steps involved. In a later section we shall draw attention to a method by which, when the apparatus is finally in use, the accuracy of the focusing can be tested.

The circle of focus and the instrument circle are to be made to coincide. The first operation is to ensure that slit S_1, after being set vertical, is at a distance $\frac{1}{2}r$ from Q, where r is the radius of curvature of the grating. This is done by direct measurement with a cathetometer device.

The next step is to make the radius of the 'instrument circle' equal to $\frac{1}{2}r$. To this end the position of S_1 is first located in space by two microscopes which are focused upon it from widely different directions. Slit S_1 is then removed upon its mounting plate P_1, and the slit S_2 is brought round by rotation of the arm A and is adjusted until it is in the position and orientation previously occupied by S_1 in the fields of view of the microscopes. The reading of the micrometer having been taken, the arm A is moved back and the slit S_1 replaced. The instrument circle now has the required radius, the slit S_1 lies upon it in a measured position and the slits are parallel to one another.

The pole of the grating must now be placed on the instrument circle at the position corresponding to the chosen angle of incidence of $87\frac{1}{2}°$. The required position is first established by placing there the slit S_2, the arm A being rotated until the micrometer reading has the appropriate value. The position is then located in space by focusing a microscope upon S_2, which is then moved back. The grating table, with the grating removed, and with its axis determined in the usual way and indicated by a vertical hair, is then moved until the image of the hair is in the position in the field of view of the microscope previously occupied by S_2. Then, the hair having been removed, the grating is mounted on the table and adjusted until its pole occupies the same position. The pole of the grating now lies on the axis of rotation of the table and both lie on the instrument circle.

The remaining adjustments (the orientation of the grating surface and of the rulings) present no special difficulties.

§ 4. The Geiger Counter

The counter consists of a vertical anode, of 0·1 mm diameter tungsten wire, and a coaxial copper cylinder of length 2 cm and diameter 1 cm. A wide vertical slit, extending along the whole length of the cylinder, allows the radiation to enter. The counter is enclosed in a glass bulb which is provided with a window, strong enough to withstand the gas pressure of several centimetres of mercury in the counter, and yet thin enough to permit an appreciable proportion of the x-rays to enter. This window consists of two superposed films of cellulose nitrate, each 500 Å thick, covering a vertical slit. The transmission of soft x-rays by such a window, interpolated from the experimental work on celluloid by Neufeldt (1931) and by O'Bryan (1932), is as follows:

Wavelength of x-rays (Å)	46	89	176	220	308	400
Percentage transmission	94	71	25	17	4	0·6

The films are formed in the usual way by floating on water and are collected successively on a brass disc, through which is a slit 5 mm long and 0·25 mm wide. They are then imprisoned between this disc and a similar one possessing a slightly wider slit. Windows formed in this manner are found to have leaks of less than 0·1 micron-litres per second for a pressure difference of 6 cm Hg. Such leaks are acceptable when the counter is connected to a ballast-volume of 2½ litres.

Although these windows are actually capable of withstanding pressures considerably greater than 10 cm Hg, they would be broken by a pressure of one atmosphere. Hence, whenever the vacuum box is opened to the atmosphere, it is necessary simultaneously to let air into the counter. For this reason, and because of the leak through the window, the counter cannot be permanently sealed off. Therefore, to avoid the difficulty of obtaining at all times a gas mixture of the same consitution, we have refrained from using a quenching gas and employed instead a simple gas, usually xenon, in the counter. Quenching is achieved by the use of a modified form of Neher–Harper circuit. As a result of the necessarily unorthodox form of the counter, the plateau is rather poor, having a length of about 30 volts and a slope of 1½% per volt.

Despite a well stabilized high tension, the counting rate was found to drift seriously, especially during the first few hours after the counter had been filled. It was found that this was caused by a change in the characteristics of the counter which resulted in an increase in the threshold voltage of the counter by as much as 20 or 30 volts. The drift in the counting rate has been prevented by employing a method developed by Brachet (1950) which enables the high tension voltage to be varied in such a way that the operating point is unchanged in relation to the plateau. Accordingly, in addition to the normal counting channel, which records all but the very small pulses from the Geiger counter, a channel is provided to count only the very large pulses. The size of pulse selected by this second channel is adjusted so that the ratio of the two counting rates is 5 : 1 when the counter is operating at the desired point on the plateau. But the proportion of large pulses depends markedly on the excess of the applied voltage over the threshold voltage, i.e. on the overvoltage. Hence, if the ratio of 5 : 1 is maintained by manual adjustment of the applied voltage, the working position on the plateau should remain constant and the drift should be avoided. In fact, the residual drift is found to be less than 2% over a period of four hours.

The finite dead time of the counter makes it unsound to assume that the counting rate is exactly proportional to the intensity of the radiation. Tests with γ-rays showed that there is a progressively increasing deviation from proportionality for counting rates greater than 3000 counts/min, and allowed a correction curve to be obtained. We have found (1951) that the corrected counting rate gives a reliable measure of the intensity of soft X-rays.

§ 5. The X-Ray Tube

The X-ray tube is of metal with water-cooled case and anti-cathode. It has no novel features of special advantage and will not be described in detail. It is, however, necessary to mention that the anti-cathode is fitted with a metal bellows device by means of which its position may be adjusted without breaking the vacuum. The tube is evacuated by an oil diffusion pump through a cold trap to a pressure which is usually about 10^{-5} mm Hg. A separate diffusion pump evacuates the spectrograph to a pressure of about 10^{-4} mm Hg.

With the filament current supplied by batteries, and with the high tension suitably stabilized, the X-ray tube gives a stable radiation intensity provided that the surface conditions of the anti-cathode remain constant. In practice this was achieved only when the anti-cathode material, in the form of a sheet, was attached to the anti-cathode block in such a way as to run hot.

§ 6. The Focus, Wavelength Calibration and Resolution of the Instrument

There is a simple and reliable method of verifying that the spectrometer is correctly focused. It may be applied quickly and without disturbance to the apparatus. The slit S_2 is set to receive the peak radiation of an emission line. If the instrument is out of focus, S_2 is not at the intersection of rays from all parts of the grating surface. Consequently its setting, for the line peak, is dependent on the part of the grating from which the radiation is being reflected. Hence, unless the spectrometer is perfectly focused, a movement of the anti-cathode in a horizontal direction across the slit S_1, causing first one edge and then the other of the grating to be irradiated, must produce a shift of the apparent wavelength of the peak. Observation of the magnitude and direction of this shift usually makes it possible not only to detect but to evaluate, and hence subsequently to correct, an error of focus.

The original adjustment of the grating permits the calculation of the angle of incidence of the radiation and makes it possible, if the constant of the grating is known, to calculate the wavelength corresponding to any micrometer reading of the position of the main arm. Corrections to these wavelength calculations have been obtained by observations of the first, second and fourth orders of the M_ξ ($M_{IV,V}$–$N_{II,III}$) line of molybdenum, for which the wavelength of 64·35 Å (Siegbahn and Magnusson 1934) has been assumed.

An estimate of the resolving power which can be expected of the instrument, when it is perfectly adjusted and when S_2 is of zero width, can be obtained by employing the expression $R = 0{\cdot}91\, r\lambda m/S\sigma$ obtained by Mack, Stehn and Edlén (1932). In this expression R is the resolving power, r the radius of curvature of the grating, m the order of the spectrum, S the width of the slit S_1 and σ the grating

constant. The expression is valid when, as in the present apparatus, the resolution is controlled by the appreciable width of the slit S_1. With the insertion of the appropriate values of the constants of the apparatus the expression becomes $R = 5{\cdot}2 \times 10^8 m\lambda$. From this result, and because an extended form of Rayleigh criterion was employed in deriving the quoted expression, the apparent line-width at half maximum intensity for a truly monochromatic radiation is a little less than $1/5 \times 10^8 m$ cm or $1/5m$ ångström units. This is the width $d\lambda$ of the line that would fall on the slit S_2; and this slit has actually a finite width corresponding to a wavelength range $d\lambda'$. It can easily be shown that $d\lambda'$, like $d\lambda$, is independent of the wavelength concerned and proportional to $1/m$. Consequently the total 'instrumental line-width' is α/m ångström units, where α, for the actual width of S_2, should now be a little greater than 0·2. There is no need to attempt a precise computation of α since, as will be shown, the effective value can be found experimentally. The observed value, 0·32, is sufficiently close to 0·2 to show that the focus is satisfactory.

§ 7. The Instrumental Line-Width and the Form of X-Ray Lines

To illustrate the behaviour and one of the uses of the instrument, the results of a few experiments on the shape of soft x-ray lines will be discussed. The lines examined are not associated with valence electronic bands and are, therefore, relatively narrow and provide a more stringent test of the apparatus. In the soft x-ray region such lines have received little quantitative attention. Their study requires the knowledge of the width and approximate shape of the 'instrumental line' that would be obtained with strictly monochromatic radiation. Their study also makes it possible to obtain this information. In the present pioneer investigation with the instrument we have elected to study lines from anti-cathode materials which do not deteriorate with time and which can be maintained hot so as to avoid the formation on their surface of contaminating films.

Weisskopf and Wigner (1930) give a theoretical expression for the intensity distribution in an x-ray line when there is no effect of interatomic interaction. Since, in the present work, we are concerned with the number of photons entering slit S_2, we write the expression in the convenient form*

$$N_{AB}(\nu)\,d\nu = \frac{\Gamma_A + \Gamma_B}{2\pi} \frac{N_0\,d\nu}{(\nu_{AB} - \nu)^2 + \{\frac{1}{2}(\Gamma_A + \Gamma_B)\}^2} \qquad \ldots\ldots(1)$$

where N_0 is the total number of photons arising from transitions between state A and state B, $2\pi\,\Gamma_A$ and $2\pi\Gamma_B$ the reciprocal mean lives of these states, and $N_{AB}(\nu)\,d\nu$ the number of photons emitted in the frequency range $d\nu$. The line should evidently be symmetrical and its frequency width at half maximum equal to $\Gamma_A + \Gamma_B$. But the form of the line actually observed depends on the form of the instrumental line. However, a simple treatment, similar to that used by Richtmyer and Barnes (1934) for the double crystal spectrometer, shows that if both the true x-ray line and the instrumental line happen to be of the form of equation (1), then the resulting observed line must also be of this shape and of width equal to the sum of the widths of the true and instrumental lines. The results will indicate that this is approximately true in the present experiments.

* From the expression $I_{AB}(\nu)\,d\nu = \dfrac{\nu}{2\pi\nu_{AB}} \dfrac{(\Gamma_A + \Gamma_B)\,d\nu}{(\nu_{AB} - \nu)^2 + \{\frac{1}{2}(\Gamma_A + \Gamma_B)\}^2}$.

Experiments on the first order of the M_ζ line of molybdenum were described in our earlier note (Rogers and Chalklin 1951). For convenience the experimental curve has been reproduced in figure 2(*a*). Its width at half maximum is 0·80 Å, and when the counting rates are converted to $N(\nu)$ and plotted against electron volts the results are found to be well fitted by a curve of the form of equation (1). This conversion has actually been made in figure 2(*b*), which represents the fourth order of the M_ζ line. Since the width of slit S_2 corresponds to a wavelength range $d\lambda$ which is independent of λ the observed counting rates are proportional to $N(\lambda)$. To convert them to $N(\nu)$ it is therefore necessary to multiply them by $d\lambda/d\nu$. The points plotted are the means of at least two counts, each of one minute duration. The counting rates due to the background of continuous and scattered radiation have been obtained by interpolation from readings on either side of the line and have been subtracted from the actual observations. The two sets of points represented in the figure were taken with slightly different x-ray tube currents, and one of them has therefore been multiplied by an appropriate factor.

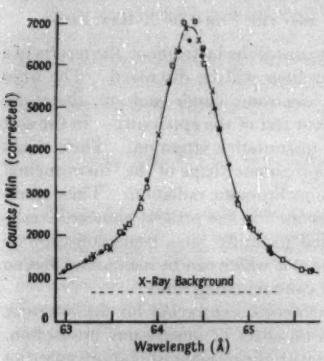


Figure 2 (*a*). Molybdenum M_ζ (first order).

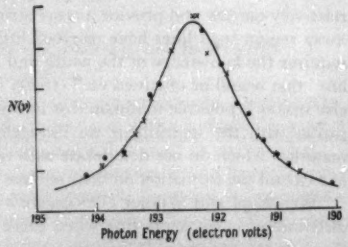


Figure 2 (*b*). Molybdenum M_ζ (fourth order).

The results are less impressive than those for the first order, for the counting rates were about 20 times smaller and the standard deviation is therefore more serious. Nevertheless they define the form of the line fairly well, and are well represented by the curve, which was drawn from equation (1) with $\Gamma_A + \Gamma_B$ equal to 1·68 ev. This width is equivalent to 0·56 Å and is much less than that of the first order. The magnitude of the difference suggests that the instrumental line-width has little influence on the observed width of the fourth order line but has an appreciable effect in the first order. Since both first and fourth order lines are of Weisskopf–Wigner form, it may be inferred that the true M_ζ line and, approximately, the instrumental line, are individually of this shape. Consequently, employing the principle of additional widths mentioned above, and expressing the widths for convenience in ångström units, $W_M + \alpha = 0{\cdot}80$ and $W_M + \frac{1}{4}\alpha = 0{\cdot}56$, where W_M is the true width of the M_ζ line and α/m the instrumental line width. Hence $W_M = 0{\cdot}48$ Å and $\alpha = 0{\cdot}32$ Å. Even in the first order α/m is less than W_M. Thus the shape and width of the instrumental line are less well determined than the shape and width of M_ζ

itself. But their accuracy is certainly sufficient to allow them to be used in the correction of other x-ray lines, provided that these lines are not narrower than M_ζ. Finally the true M_ζ line, although composite and represented formally by $M_{IV,V} - N_{II,III}$, is seen to be symmetrical and approximately of Weisskopf–Wigner form, with a width at half maximum of 0·48 Å or 1·44 ev.

§ 8. Application to Spectrum of Tungsten

Results for the N spectrum of tungsten are shown in figure 3, in which the scale of equivalent electron volts has been slightly adjusted* so that the two main emission peaks shall agree with the standard values, 58·5 Å (211·6 ev) and 55·8 Å (221·8 ev), obtained by Magnusson (1932). The interpolated background counting rates have been subtracted from the observed values, the interpolations being performed with the help of observations, not shown in figure 3, on either side of the group of emission lines. The background is relatively strong and contributes appreciably to the random variation of the counts.

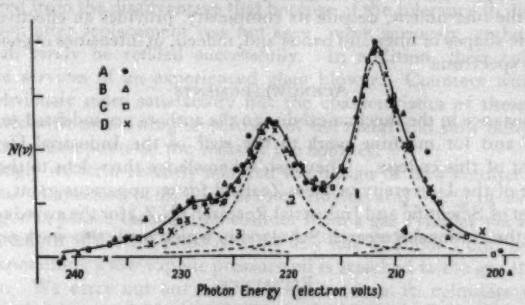


Figure 3. Analysis of the tungsten emission 'doublet' $N_{IV,V}$–$N_{VI,VII}$. Tube voltage: A, B and C, 4000; D, 900. Maximum counts per minute (approx.): A, 2000; B, 3500; C, 7500; D, 4000.

Although obtained with widely different tube currents, runs A, B and C agree well. The spectrum is more complicated than had been supposed, and, indeed, four lines of Weisskopf–Wigner form are needed to interpret the experimental results. The full line in the figure, which fits the experimental results satisfactorily, is the resultant of the component lines marked 1, 2, 3 and 4. The characteristics of these lines are set out in the table:

Line	1	2	3	4
Interpretation:	$N_V - N_{VII}$	$N_{IV} - N_{VI}$	$N_{II} - N_{IV}$	$N_V - N_{VI}$
Wavelength (Å)	58·50	55·84	54·07	59·58
Relative intensity (in no. of photons observed)	1	0·73	see below	~0·09
Burger–Dorgelo rule	1	0·81	—	0·05
Corrected width (ev)	3·9	4·7	4·7	—

* The adjustment actually consisted of setting line 1 (figure 3) at 58·50 Å.

The agreement of the relative intensities of lines 1 and 2 with the Burger–Dorgelo rule (including the ν^3 term) is satisfactory, but although the intensity of line 4 is of the order of magnitude predicted for N_V–N_{VI}, the general significance of this effect must be considered doubtful.

The shoulder on the curve, associated with line 3, appears to be weaker in run D than in the runs at 4000 volts. Its excitation voltage, although less than 900, is therefore presumably greater than that of lines 1 and 2 (about 250 volts). The N_I, N_{II} and N_{III} levels occur in this range. We actually attribute line 3 to N_{II}–N_{IV}, whose wavelength, calculated from appropriate L lines, should be about 52·7 Å (234·7 volts). This is sufficiently close to the observed value of 54·07 Å.

Some satellite structure may of course be present, undisclosed by the results of figure 3 and by the adopted analysis. This point is being investigated in this laboratory, but it is presumed that the general outline of the experimental curve is essentially due to the diagram lines considered.

These results are the first obtained with this type of instrument and there is no doubt that improvements can be made in our techniques. But the experiments show that the instrument, despite its complexity, provides an effective means of studying the shapes of lines and bands and, indeed, of intensities in general in this part of the spectrum.

Acknowledgments

For assistance in the mechanical design the authors are indebted to Mr. F. C. Gallagher, and for machine work to the staff of the Industrial Development Department of this college. They also acknowledge their debt to the Research Committee of the University of New Zealand for an apparatus grant, and to the Department of Scientific and Industrial Research (N.Z.) for the award to one of us (J.L.R.) of the National Research Scholarship which made this work possible.

References

Brachet, C., 1950, *C. R. Acad. Sci., Paris*, **230**, 2174.
Mack, J. E., Stehn, J. R., and Edlén, B., 1932, *J. Opt. Soc. Amer.*, **22**, 245.
Magnusson, T., 1932, *Z. Phys.*, **79**, 161; 1938, *Nova Acta Reg. Soc. Sci., Upsaliensis*, Ser. IV, Vol. II, No. 3.
Neufeldt, H., 1931, *Z. Phys.*, **68**, 659.
O'Bryan, H. M., 1932, *J. Opt. Soc. Amer.*, **22**, 739.
Piore, E. R., Kingston, R. H., Gyorgy, E. M., and Harvey, G. G., 1951, *Rev. Sci. Instrum.*, **22**, 543; 1952, *Ibid.*, **23**, 8.
Richtmyer, F. K., and Barnes, S. W., 1934, *Phys. Rev.*, **46**, 352.
Rogers, J. L., and Chalklin, F. C., 1951, *Proc. Phys. Soc.* A, **64**, 935.
Siegbahn, M., and Magnusson, T., 1934, *Z. Phys.*, **88**, 559.
Weisskopf, V., and Wigner, E., 1930, *Z. Phys.*, **63**, 54.

RESEARCH NOTES

A Gas-Flow X-Ray Diffraction Counter

By U. W. ARNDT, W. A. COATES and A. R. CRATHORN
The Royal Institution, London W.1

Communicated by D. P. Riley; MS. received 3rd December 1953

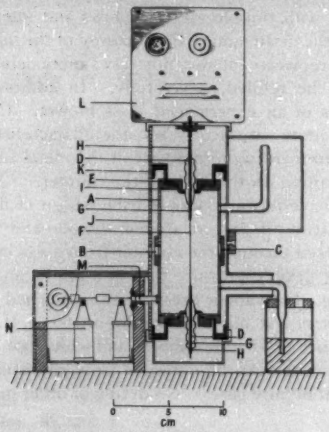
The use of proportional counters for X-ray diffraction purposes has been described in a number of communications (Lang 1951, 1952, Arndt and Riley 1952, Arndt, Coates and Riley 1953). Our own earlier counters had glass envelopes with bubble glass windows and internal carbon cathodes. They suffered from the disadvantage that because of the tolerance in dimensions, exact location after replacement was not easy; more seriously, carbon cathode counters can rarely be refilled successfully. In addition, their construction requires the services of an experienced glass blower. Counters with a metal body are obviously more satisfactory but the characteristics of these counters change slowly with time owing to porosity of the metal and slow out-gassing of or, alternately, absorption on the walls of the counter.

Accordingly we have recently adopted the design of flow counter shown in the diagram. A mixture of $2\frac{1}{2}\%$ of carbon dioxide and $97\frac{1}{2}\%$ of 99·8% pure cylinder argon is passed through the counter from a gas cylinder via a reducing valve and a length of flexible pressure tube at a rate of about 1 cm^3 per second. A bubbler containing a low vapour pressure oil is attached to the gas exit tube of the counter. We carry out our own mixing in 60 cu. ft. cylinders which are filled to a total pressure of 75 atmospheres and which last for about one month. If preferred, gas mixtures can be obtained from the usual suppliers at a cost little greater than that of pure gases. No drying or other purification of the gases appears necessary.

The body of the counter A is of copper and the incident radiation enters through a window B. For copper characteristic radiation the window is made of nickel and serves as a β-filter. Any radiation not absorbed by the gas is trapped by a lead backstop C to prevent the generation of fluorescent radiation. In order to reduce photoelectric effects in the counter the copper cathode is rendered passive by the procedure recommended by Curran and Craggs (1949). To facilitate this chemical treatment and the subsequent washing, the glass–metal terminals D are soft soldered into brass plugs E. All the soldering on the main body is carried out first and this is then pickled and washed. The 0·002 inch tungsten wire F is spot-welded to short nickel rods G and after these have been passed through the central tubes H of the terminals the plugs are screwed to the end plates I of the counter. O-rings are used to make gas-tight connections. The nickel rods are soft-soldered to the tube while the wire is being stretched by a load. The wire is glowed in an argon atmosphere before use. The counter is held in its screening can J by means of 'Bakelite' insulators K locating in the end-plates I and the central wire is attached directly to the grid

of the first valve of the pre-amplifier L. The negative high tension connection M is taken to the case of the counter through a filter N, and an earth connection is made to the screening cans by means of one cable only in order to prevent loop pick-up, the counter and its support being insulated from the counter table.

The construction and assembly are very simple and require no glass-blowing or instrument-making facilities; the glass–metal seals are commercially available (Messrs. Cathodeon Ltd., Type 3008) with a central tubular conductor instead of the usual rod. It should be noted that the incident beam is well collimated and the ionization within the argon is thus confined to a short section of the counter. There is thus no need for the counting volume to be limited with great accuracy and guard-ring electrodes are unnecessary. With the gas mixture referred to above the operating potential of the counter is about 1800 volts. Under reasonably dry atmospheric conditions there is no indication of electrical breakdown across the electrodes until a much higher voltage is reached.



The detection efficiency of the counter for CuKα radiation is about 70% and no variation with atmospheric pressure or with time could be detected. The fluctuation in the pulse size for CuKα quanta is given by $\sigma^2/\bar{P}^2 = 0{\cdot}0034$, where $\bar{P}$ is the average pulse size and σ the root mean square deviation of this size. This fluctuation is somewhat smaller than previously reported values (Curran, Cockroft and Angus 1949, Lang 1952, Arndt, Coates and Riley 1953). The counter is used with a linear amplifier type 1008, a single channel pulse analyser type 1168 and a power supply type 1007.* The gas amplification used is about 100 and the external electronic amplification 10^5. Under these conditions the mean pulse height was found to be subject to random drifts of $\pm 3\%$ for any given filling of the gas cylinder (the greater part of this variation can be attributed to the circuits). When the counter is in use for measuring diffraction spectra the channel is set to accept pulses within $\pm 25\%$ of the mean pulse height and the drifts are therefore of no consequence. In this manner of operation the argon

* Atomic Energy Research Establishment, Harwell, type numbers.

fluorescence peak is necessarily lost. The amplifier is operated with differentiating and integrating time constants of 0·6 μsec so that very fast counting is possible.

These counters were developed as part of a research programme supported by the British Empire Cancer Campaign.

REFERENCES

ARNDT, U. W., COATES, W. A., and RILEY, D. P., 1953, *Proc. Phys. Soc.* B, **66**, 1009.
ARNDT, U. W., and RILEY, D. P., 1952, *Proc. Phys. Soc.* A, **65**, 74.
CURRAN, S. C., COCKROFT, A. L., and ANGUS, J., 1949, *Phil. Mag.*, **40**, 929.
CURRAN, S. C., and CRAGGS, J. D., 1949, *Counting Tubes* (London : Butterworths Scientific Publications), p. 116.
LANG, A. R., 1951, *Nature, Lond.*, **168**, 907 ; 1952, *Proc. Phys. Soc.* A, **65**, 372.

Interferometric Observation of Mosaic Structure on the (111) Face of a Single Crystal of Germanium

BY AJIT RAM VERMA*
Royal Holloway College, University of London, Englefield Green, Surrey

MS. received 25th November 1953

RECENTLY Avery and Clegg (1953) have measured the optical constants of a highly perfect single crystal of germanium grown from the melt at Marconi Wireless and Telegraph Co. Limited, Chelmsford. It was therefore thought to be of interest to study the surface of the same crystal by interferometric techniques, and the author is indebted to Dr. I. Cressell for making available the crystal for this study.

The ingot of this germanium crystal has an exceptionally well developed large sized (nearly 5 mm across) hexagonal-shaped (111) face which appears quite smooth and plane when viewed under a metallurgical microscope using bright field illumination. This natural (111) face, without any polishing or chemical etching, was therefore examined by a vertical positive phase-contrast microscope. The appearance of the surface is shown in figure 1†, where the entire crystal face is seen to be covered with a mosaic structure, the figures being hexagonal in shape, varying in size, and also observed sometimes to be arranged along a row. To study these features interferometrically, the crystal surface was matched against a silvered glass flat. However, to save the crystal face from subsequent cleaning by hydrogen peroxide, which may etch the surface, it was not silvered. The multiple-beam interference fringes so obtained are, however, fairly sharp, due to the high reflectivity of the germanium crystal. Figure 2 shows the multiple-beam Fizeau fringes passing over the crystal face. Due to the high sensitivity of these fringes to reveal the surface topography, the mosaic structure is clearly visible within the width of the fringes. In order to decide whether these features are elevations or depressions over the general level of the crystal, and to get the magnitude of the height or depth, a line section was selected by a spectrograph slit and the fringes of equal chromatic order (Tolansky 1948) were utilized. Figure 3 shows such a line section, which was adjusted so that at

* I.C.I. Research Fellow.

† Figures 1, 2, 3 are printed in Plate, at end of issue.

its middle part it passed over one of the deepest hexagonal mosaic figures. The V-shaped fringes, convex towards the red end of the spectrum, indicate the hexagonal figure to be a depression, and in this particular case the depth is approximately 900 Å. Other shallower features over which the line section passes give the fringe a wavy appearance. These depressions are not flat-bottomed but are formed by sloping faces; in the above case the slope is about 10′. Indeed, in the micrograph (figure 1) it can be seen that these hexagonal figures have a structure with a point at the centre from whence the sloping surfaces radiate.

The mosaic boundaries can be interpreted in terms of dislocations as a row or lattice of dislocations as was first proposed by Burgers (1940). Evidence that the lineage boundaries consist of a row of dislocations has recently been produced by Vogel *et al.* (1953) who studied the boundaries between nearly perfect crystals of germanium grown from the melt, the growth direction being ⟨110⟩. But the crystal had to be etched in a suitable reagent to reveal the end points of the edge dislocations which serve as nuclei for etch pits (Lacombe and Beaujard 1947). However, it is possible to see the mosaic structure on the natural surfaces of a highly perfect germanium crystal by interferometric techniques without the need for etching the surface.

ACKNOWLEDGMENT

The author wishes to thank Professor S. Tolansky for the facilities of the laboratory and his interest.

REFERENCES

AVERY, D. G., and CLEGG, P. L., 1953, *Proc. Phys. Soc.* B, **66**, 513.
BURGERS, J. M., 1940, *Proc. Phys. Soc.*, **52**, 23.
LACOMBE, P., and BEAUJARD, L., 1947, *J. Inst. Metals*, **74**, 1.
TOLANSKY, S., 1948, *Multiple Beam Interferometry of Surfaces and Films* (Oxford: University Press).
VOGEL, F. L., PFANN, W. G., COREY, H. E., and THOMAS, E. E., 1953, *Phys. Rev.*, **90**, 489.

On the Thermal and Electrical Conductivity of Semiconductors

BY H. J. GOLDSMID
Research Laboratories of The General Electric Company, Limited, Wembley, England

MS. received 23*rd December* 1953

THE ratio of electrical conductivity to thermal conductivity is approximately constant for all metals at the same temperature owing to the fact that the free electrons are responsible for carrying both the electrical and thermal currents, this being the long established Wiedemann–Franz law. In the discussion which follows the semiconductors are assumed to be of such purity that there are very few electrons available to contribute to the flow of heat. In such circumstances the thermal current may be considered to be carried solely by elastic waves, as is the case for electrical insulators. The thermal conductivity is then governed entirely by the crystal structure, and it might not be thought that it should bear any relation to the electrical conductivity which is, of course, still dependent upon the free electrons. However, there is one property which is directly related to the electrical conductivity, and which, at the same time, depends on the crystal structure, this property being the mobility of the electrons. It is therefore proposed to consider whether or not there is any relation between mobility and thermal conductivity in semiconductors.

The mobility of electrons in a semiconductor is limited by two forms of scattering, firstly by impurity centres (Conwell and Weisskopf 1946, 1950), and secondly by the thermal vibrations of the lattice. For the high purity semiconductors, which are dealt with here, it will be assumed that the scattering is entirely due to the vibrations of the lattice. Shockley and Bardeen (1950) have discussed such scattering in terms of deformation potentials. The boundaries of the energy bands of a crystal are dependent on its lattice parameter, and owing to thermal vibrations the latter undergoes transient changes about its mean value. An electron moving through the lattice is therefore subject to changes of kinetic energy, and for certain components of the thermal waves these changes of energy give rise to the possibility of reflection. There is thus a scattering effect which is found to be random in direction. The mobility μ is given as

$$\mu = \tfrac{4}{3}\boldsymbol{e}l_e(2\pi \boldsymbol{m}\boldsymbol{k}T)^{-1/2} \qquad \ldots\ldots(1)$$

where l_e is the mean free path of electrons, $\boldsymbol{e}$ the electronic charge, $\boldsymbol{m}$ the mass of the electron, $\boldsymbol{k}$ Boltzmann's constant and T the absolute temperature.

The present state of the theory of thermal conductivity in dielectric crystals is largely due to Peierls (1929), who refined Debye's original theory, taking into account the lattice structure. The elastic waves which are responsible for the flow of heat may be quantized as phonons, and in a large crystal, free from defects or impurities, the thermal resistance arises from collisions between the phonons themselves. At very low temperatures scattering from the boundaries of the crystal must be considered, but at all higher temperatures the important processes are the 'Umklapp' processes which lead to random reflections. The thermal conductivity λ is given as

$$\lambda = \tfrac{1}{3}l_t c v/\rho \qquad \ldots\ldots(2)$$

where l_t is the mean free path of phonons, c the specific heat, ρ density and v the velocity of sound. The velocity of sound may be derived from the elastic constants of the material.

Using equations (1) and (2), the mean free paths of electrons and phonons may be calculated from the electron mobility and the thermal conductivity respectively. It is interesting to compare these mean free paths since in both cases there is random scattering. Unfortunately there are very few semiconductors where all the required properties have been measured, but five such materials have been found and the results are shown in table 1. All the measurements are quoted as having been made on large crystals as opposed to powdered specimens, and in each case the highest value of the mobility is used, it being assumed that this corresponds to the limitations imposed by the crystal lattice only. It is remarkable that in every case the mean free path of both electrons and phonons is of the same order.

From (1) and (2)

$$\frac{\mu}{\lambda} = \frac{4\boldsymbol{e}\rho}{cv(2\pi \boldsymbol{m}\boldsymbol{k}T)^{1/2}}\frac{l_e}{l_t}. \qquad \ldots\ldots(3)$$

In the light of the results of table 1 it will be assumed that as an approximation l_e and l_t can be equated. The following expression is then obtained:

$$\frac{\mu}{\lambda} \simeq K\frac{\rho}{cv}, \qquad \ldots\ldots(4)$$

K being a constant for all materials at the same temperature.

Table 1. The Thermal and Electrical Conductivity of Semiconductors

	ρ (gm cm^{-3})	c (cal gm^{-1} deg^{-1})	λ (cal cm^{-1} sec^{-1} deg^{-1})	v (10^5 cm sec^{-1})	μ (cm^2 v^{-1} sec^{-1})	l_t (cm $\times 10^{-6}$)	l_e (cm $\times 10^{-6}$)	l_t/l_e
Carbon (diamond)	3·51	0·120	0·36[a]	17·70[d]	900[i] (Hall)	17·8	6·4	2·8
Silicon	2·33	0·170	0·20[b]	9·16[e]	1200[j] (Drift)	9·0	8·6	1·05
Germanium	5·33	0·074	0·12[c]	5·41[f]	3600[k] (Drift)	48·0	25·6	1·9
Lead sulphide	7·5	0·050	0·0016[a]	3·43[g]	640[l] (Drift)	2·1	4·6	0·45
Tellurium	6·25	0·048	0·015[a]	2·84[h]	1600[m]	20·6	11·4	1·8

[a] *International Critical Tables.*
[b] *Metals Handbook.* Amer. Soc. for Metals (1948).
[c] Measurement by author.
[d] Bhagavantam and Bhimasenachar, *Proc. Roy. Soc.* A, 1946, **187,** 381.
[e] Bridgman, *Proc. Amer. Acad. Arts and Sci.*, 1949, **77,** 187.
[f] Bond *et al.*, *Phys. Rev.*, 1950, **78,** 176.
[g] Ramachandran and Wooster, *Nature, Lond.*, 1949, **164,** 839.
[h] Shockley and Bardeen, *Phys. Rev.*, 1950, **77,** 407 (from compressibility).
[i] Klick and Maurer, *Phys. Rev.*, 1949, **76,** 179.
[j] Haynes and Westphal, *Phys. Rev.*, 1952, **85,** 680.
[k] Haynes and Shockley, *Phys. Rev.*, 1951, **81,** 835.
[l] Putley, *Proc. Phys. Soc.* B, 1952, **65,** 388.
[m] Fukuroi *et al.*, *Sci. Rep. Res. Inst. Tôhoku Univ.*, 1949, **A1,** 365, 375, 1950, **A2,** 233, 239.

Table 2. The Thermal and Electrical Conductivity of Semiconductors

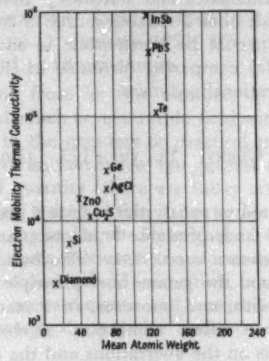
	A	λ (cal cm^{-1} sec^{-1} deg^{-1})	μ (cm^2 v^{-1} sec^{-1})
AgCl	71·7	0·0026[n]	50[p] (Hall and Drift)
Cu_2S	53·1	0·0011[n]	12[q] (Hall)
ZnO	40·7	0·0091[n]	100–200[r] (Hall)
InSb	118·3	0·04[o]	38000[s]

[n] *International Critical Tables.*
[o] Measurement by author.
[p] Haynes and Shockley, *Phys. Rev.*, 1951, **82,** 935.
[q] Hirahara, *J. Phys. Soc. Japan*, 1951, **6,** 428.
[r] Hahn *et al.*, *Phys. Rev.*, 1949, **75,** 1631.
[s] Gremmelmaier and Madelung, *Z. Naturforsch.*, 1953, **8a,** 333.

Further, from Kopp and Neumann's extension of Dulong and Petit's law, cA is approximately constant for all materials where A is the mean atomic weight, which is defined as the molecular weight divided by the number of atoms in the molecule. Then from (4)

$$\frac{\mu}{\lambda} \simeq K' \frac{\rho A}{v} \qquad \ldots\ldots(5)$$

where K' is another constant.



In general it may be stated that as the atomic weight increases the density increases and the velocity of sound decreases. Thus the ratio μ/λ should become greater the larger the mean atomic weight. Table 2 shows the mobility and thermal conductivity for four additional materials, and the diagram illustrates the rapid rise of μ/λ with A. This is most remarkable in view of the fact that materials of quite widely different crystal structures have been included. It is interesting to compare InSb and PbS, which have almost the same mean atomic weight. Even though the mobility in the former is some sixty times that in the latter, the ratios μ/λ differ only by a factor of about 2.

Since writing this note the attention of the author has been drawn to a paper by Joffe (1952) on the estimation of thermal conductivity of semiconductors. In this paper mention is made of the fact that the thermal conductivity of an element is related to its position in the Periodic Table, but no comparison between electrical and thermal conductivity is made, it being assumed that conditions of scattering for phonons are different from those for electrons in semiconductors.

Acknowledgments

The author would like to thank Mr. R. W. Douglas for many helpful discussions.

References

Conwell, E., and Weisskopf, V. F., 1946, *Phys. Rev.*, **69**, 258 ; 1950, *Ibid.*, **77**, 388.
Joffe, A. F., 1952, *Dokl. Akad. Nauk. U.S.S.R.*, **87**, 369.
Peierls, R., 1929, *Ann. Phys., Lpz.*, **3**, 1055.
Shockley, W., and Bardeen, J., 1950, *Phys. Rev.* **77**, 407.

REVIEWS OF BOOKS

Physical Formulae, by T. S. E. THOMAS. Pp. vii+118. (London: Methuen, 1953.) 8*s*. 6*d*.

A useful addition to the well-known series of Monographs on Physical Subjects, containing formulae, short explanations of principles and statements of laws in all the branches of physics encountered in an Honours degree course and also formulae in mathematics and statistics. Both graduates and undergraduates will find this little book a reliable source of reference.

In a future edition it would be worth while to add more data on M.K.S. units and to include tables comparing formulae in electricity and magnetism using unrationalized and rationalized units. J. YARWOOD.

Geometrical Optics, by C. CURRY. Pp. viii+173. (London: Edward Arnold, 1953.) £1 1*s*.

Dr. Curry has attempted to satisfy the long-felt need for a satisfactory text on geometrical optics for undergraduate students; and, it should be said at once, he has in great measure succeeded. The basic facts about reflection, refraction, lens systems, and the geometrical principles of optical instruments are all adequately dealt with, and, morever, in a manner which reduces the customary elementary, but often cumbrous, algebra to more reasonable proportions. The sections on the aberrations and the design of optical instruments, although necessarily rather bare, are a good deal better than any other treatment at this level that the reviewer can call to mind.

The labour of compiling a textbook is such that it often seems somewhat ungrateful on the part of a reviewer to criticize minor points or to express preference for a different approach. It is with this humbling thought in mind that this part of the present reviewer's task is embarked upon.

The sign-convention adopted requires that lengths be positive in the direction of propagation of the light and vice versa. When light travels from the left this leads to the sign conventions of cartesian geometry, but these are in part reversed if a reflecting surface is encountered. They are conserved, however, if the refractive index of a medium is regarded as positive for light travelling along the positive direction of the axis, and negative for light travelling in the opposite direction—which is natural enough if one recalls that the refractive index is best defined as the ratio of two velocities. In consequence this system, in which the formulae given in this book for light travelling from the left are unchanged no matter what combination of surfaces occurs, seems to have the advantage.

The diagrams are well drawn and clear, although two of them, at least, are misleading. In figure 3B, the condensing lens would in practice be biconvex, and moreover it would form an image of the lamp at or near the projection lens. Again the positive lenses in figure 7U should have almost plane surfaces facing the biconcave lens.

The account of the aberrations of lens systems is very commendably written —although some minor errors occur. For example the conditions given for the elimination of distortion are those for zero distortion for *all* object positions.

Most workers in optics nowadays would prefer to see a wave treatment of aberrations, which is both simple and avoids the fallacious conclusions that emerge from considerations of ' circles of least confusion '.

There is a useful bibliography given as an Appendix, and a short but adequate index. We may conclude that an extremely useful book on a somewhat neglected topic has appeared, for which both students and their teachers will have cause to be grateful to the author. H. H. HOPKINS.

Journal of the Audio Engineering Society, edited by L. S. GOODFRIEND. Issued quarterly (Vol. 1, no. 1, January 1953, 169 pp.). (New York: Audio Engineering Society*.) $8 per annum.

This journal, judging from its first number, appears to be a sound and well presented periodical which covers a wide field within its terms of reference. Its primary interest is to the engineer or practical physicist, and although the quality of the articles varies somewhat the general standard is good. If this standard of quality and quantity (there are thirty articles in the first number) is maintained it will satisfy a need within its particular field. This issue contains papers presented at the Audio Fair in New York in October 1952. The scope of the articles varies from network transformation to the generation of gun-shot sounds, and covers sound recording, audio amplifier design and construction problems, measurements and acoustic matters. In general the articles have a bias towards the popular presentation rather than the highly technical exposition.

Authors include some well-known names, and the following are contributors: Alan Bloch and L. S. Goodfriend of the Audio Instrument Co., J. D. Bick and W. R. Ayres of R.C.A. Victor, D. E. Maxwell of G.E.C., R. H. Tanner of Northern Electric Co., Canada, F. R. Bies of the Bell Telephone Laboratories, and W. R. Thurston of G.R. Co.

Among the articles in this number are the following: " A New Pocket Wire Recorder " by Oliver Read, which describes a battery-operated recorder of German make which is about the size of a fat ' Penguin ' novel; " Musical Therapy " by R. L. Cardinell, which is a brief review of the use of music in medicine; " Gun-Shot Reinforcers and Synthesizers " by S. L. Hathaway and R. E. Lafferty, in which the problem of the realistic, safe and accurate production of gun-shots in audio and television studios is discussed and solutions outlined.

A group of articles on sound recording include a review of binaural disc recording, an outline of methods of measuring the surface induction of magnetic tape, and a report on the characteristics and production of a new magnetic tape.

Several articles deal with the choice of component types such as resistors, electrolytic capacitors and valves for audio circuits and on the use of printed circuits.

Four articles deal with the subjective effects of non-linear distortion in audio amplifiers and disc recording equipment, and outline the advantages of different forms of measurement as a basis of assessing the performance of such equipment. By and large these authors seem to prefer the intermodulation method with the C.C.I.F. standard system favoured for some applications and the S.M.P.T.E. system for others.

* Address: P.O. Box 12, Old Chelsea Station, New York 11, N.Y., U.S.A.

"Audio-Frequency Input Circuits", by W. B. Snow, is a useful article on a much neglected subject but unfortunately there are some gaps in the general picture. There are three other articles on the practical aspects of the design of audio systems for broadcasting, television and other uses.

An article by W. E. Lehnert on "Some Factors concerning the Use of Audio Transformers" is a welcome contribution outlining the basic characteristics and their effect upon the performance of such transformers; iron distortion, longitudinal balance, magnetic shielding and high step-up ratios are dealt with.

Four theoretical articles on the use of resistance–capacity networks in amplifier design and on equalizer networks and network transformation give most useful information in these fields.

Finally a group of articles of a less precise nature deal with acoustical problems including binaural reproduction and dynamic level control in domestic apparatus and with sound-level measurements as a means of adjusting sound-reproducing equipment in domestic and other applications.

The presentation of the journal is clean and tidy. A valuable detail is the repetition of the title of the journal and its issue above the title of each article, a feature which will appeal to those who tear out or photostat pages for filing. There are 169 pages of text, and a short biography of each author is presented at the back. Only fourteen articles give a bibliography or references, while some of these are very scanty. H. D. E. ELLIS.

Fluorescence of Solutions, by E. J. BOWEN and FRANK WOKES. Pp. vii + 91. (London: Longmans, Green, 1953.) 25*s*.

This book is likely to be of most use to those engaged in biological assay work, who, after refreshing their ideas on the theoretical basis of fluorescence in liquids by the earlier chapters, will find in the last third of the book much information on the practical aspects. The subject is introduced by a simple and interesting discussion of light absorption and emission, leading to the description of potential energy curves and wave patterns in fluorescent molecules. Two chapters on quenching reach a more advanced level and are the most interesting part of the book to the physical chemist. Detection and measurement of fluorescence are then considered, with details of the calibration and operation of a number of commercial instruments. As an introduction to fluorescence, the theoretical treatment is deliberately limited and does not refer to the wide field of fluorescence in solids.

The student, to whom the book is also addressed, may well ask whether 91 moderately sized pages are worth the high price, especially when one of the plates, which presumably cost a good deal, is a photograph of a commercial instrument, available in trade catalogues, while the other is a spectrogram which could have been shown as a black and white diagram with a literature reference to the original.

A criticism which concerns all types of reader is that insufficient stress is laid on inherent weaknesses in several links in the chain of measurement, such as the instability of mercury arcs and the temperature and fatigue effects which occur in photocells. The treatment of emission cells and multipliers is somewhat inadequate, and not entirely accurate on the constitution of emissive surfaces. Another point is that all the spectrum diagrams follow the usual practice of

increasing wavelength from left to right except, for no obvious reason, figures 4 and 22. In any future edition a little valuable space could be saved by removing unnecessary repetitions, e.g. on the Raman effect, HPMV lamps, and filters.

With these exceptions the book appears to be sound and well written. The main objection to it is its price, which is excessive by comparison with other modern scientific books. S. T. HENDERSON.

Copper in Instrumentation, issued by the Copper Development Association. C.D.A. Publication No. 48. Pp. 152. (Radlett: Copper Development Association, 1953.)

"The aim of this book is to show the part copper and its alloys play in industrial instrumentation and automatic control apparatus." There is thus no secret about the intentions of this useful little survey, which pays particular attention to the range of devices where some physical or mechanical property of copper or one of its alloys offers attractive means of sensing the quantity to be measured, be it pressure, strain or temperature.

With so wide a field to cover (and the use of copper in electrical instruments is also surveyed) it is not surprising that the treatment is generally brief but the sources of more detailed information are noted in an extensive bibliography. Apart from its utility as a source book there is another direction in which the compilers are to be commended. Materials are seldom as perfect in their behaviour as the instrument designer would wish them to be. Many difficulties can, however, be avoided by selecting materials of optimum composition and in the most suitable state of cold-working or heat-treatment, and this fact, not always remembered, has been well borne in mind. G. L. J. B.

CONTENTS OF SECTION A

PAGE

Dr. B. Donovan. The Magneto-Resistance Effect in Metals at High Frequencies . 305

Dr. D. C. Bisset and Dr. J. Iball. X-Ray Diffraction from Built-up Multilayers Consisting of only a few Monolayers 315

Dr. A. C. Clark. The Binding Energy of the Alpha Particle 323

Dr. S. K. Ghosh, the late Dr. G. M. D. B. Jones and Prof. J. G. Wilson. Ionization by Relativistic μ-Mesons in Oxygen 331

Dr. A. Dalgarno and Dr. G. Poots. Approximate Molecular Orbitals: I—The $1s\sigma_g$ and $2p\sigma_u$ States of H_2^+ 343

Dr. S. Datta Majumdar. Energy Levels of Triatomic Molecules 351

Dr. R. F. Barrow, Mr. G. Drummond and Mr. P. B. Zeeman. Rotational Analysis of the β Band System of the NS Molecule 365

Research Notes:

Mr. R. A. Ballinger and Dr. N. H. March. Extended Thomas–Fermi Methods 378

Dr. E. W. Lee. The Approach to Saturation Magnetostriction 381

Dr. P. J. Price. Perturbation Theory for the One-Dimensional Wave Equation 383

Dr. J. Hatton and Dr. B. V. Rollin. The Electrical Properties of Indium Antimonide at Low Temperatures. 385

Mr. N. G. McCrum and Mr. C. A. Shiffman. The Optical Constants of Tin below the Superconducting Transition Temperature 386

Dr. D. ter Haar and Mr. A. W. Ross. Volume Changes in a Substitutional Alloy 388

Letters to the Editor:

Dr. P. J. Grant. The Coupling of Angular Momenta in the Reaction $^{27}Al(p, \gamma)^{28}Si$ 392

Mr. W. R. Hindmarsh. Spectroscopic Isotope Shifts and Electron Scattering by Nuclei 393

Corrigendum (Brodie) 394

Reviews of Books 395

Contents of Section B 396

×1140

Figure 6. Portion of interferogram of a 1 mm steel ball with scratches. Taken with Zeiss–Linnik interferometer-wedge angle varying from 21° to 26°. Fringes contour surface accurately only near the zero order fringe.

(*a*)

×1640

(*b*)

×1640

(*a*) Focused on the reference surface.
(*b*) Focused 6 wavelengths below reference surface.

Figure 7. An example of the distortion of fringe pattern due to change of lateral shift with variation of focus. Fizeau system, 2-mm oil lens, mercury green light.
The object is a grain boundary groove in cadmium. The wedge angle on either side of the groove is about 23°, so that there is a sudden change of 45° in the wedge angle at the base of the groove.

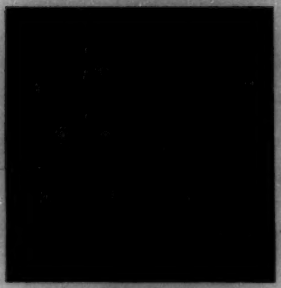


Figure 1.
Ground. 1·5 kg.
69×

Figure 2.
Ground. 2 kg.
69×

Figure 3.
Ground. 3 kg.
69×

Figure 4.
Polished.
69×

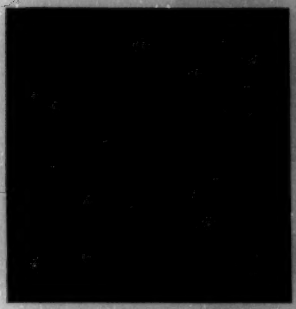


Figure 5.
Lapped. 0·25 kg.
69×

Figure 6.
Lapped. 1·5 kg.
69×

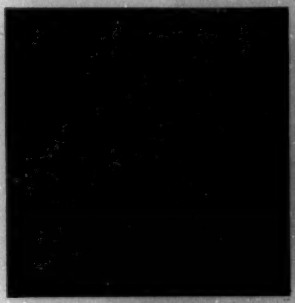


Figure 7.
Lapped. 3 kg.
69×

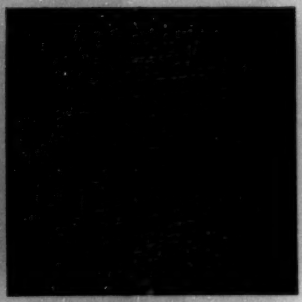


Figure 8.
Lapped and pressed.
69×

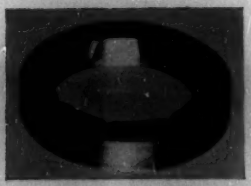


Figure 3 (*a*). Notched appendage.

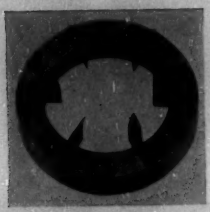


Figure 3 (*b*). Cambered radial vanes.

Figure 3 (*c*). Gauze extension.

PLATE I.

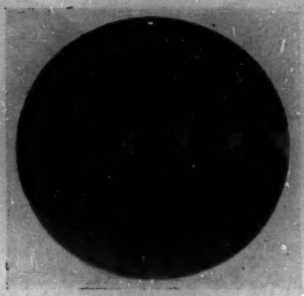


Figure 5. Flow with single notch.

Figure 6. Flow with 16 notches.

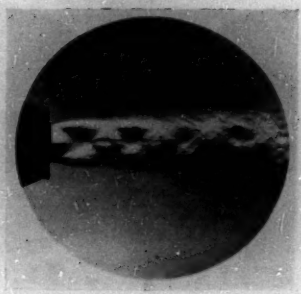


Figure 7. Flow from plain exit.

PLATE II.

Figure 14.

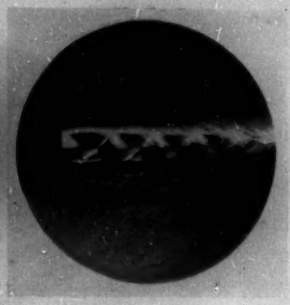


Figure 15.

Influence of a single vane.

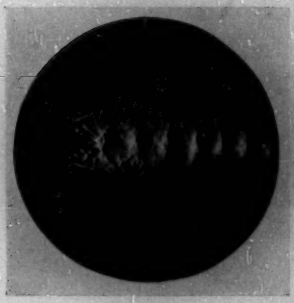


Figure 16.

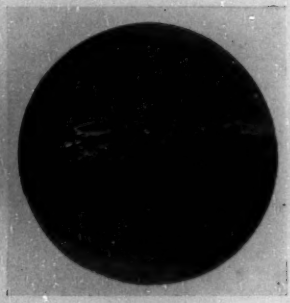


Figure 17.

Influence of six radial vanes.

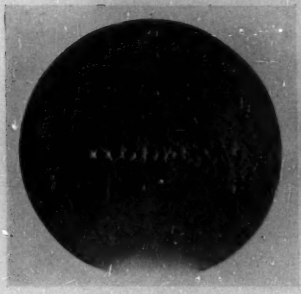


Figure 19.

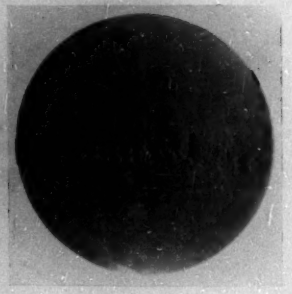


Figure 20.

Gauze extension sealed and open.

PLATE III.

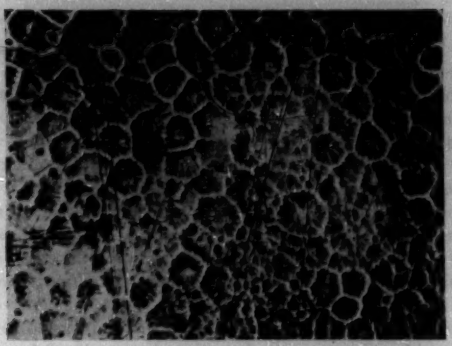


Figure 1. ×100. (*Note*. The scratches visible in the photograph are due, not to polishing, but to continuous handling of the crystal by different observers.)

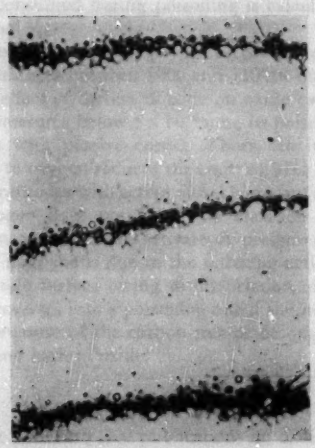


Figure 2. ×30.

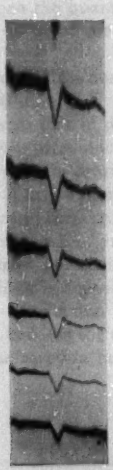
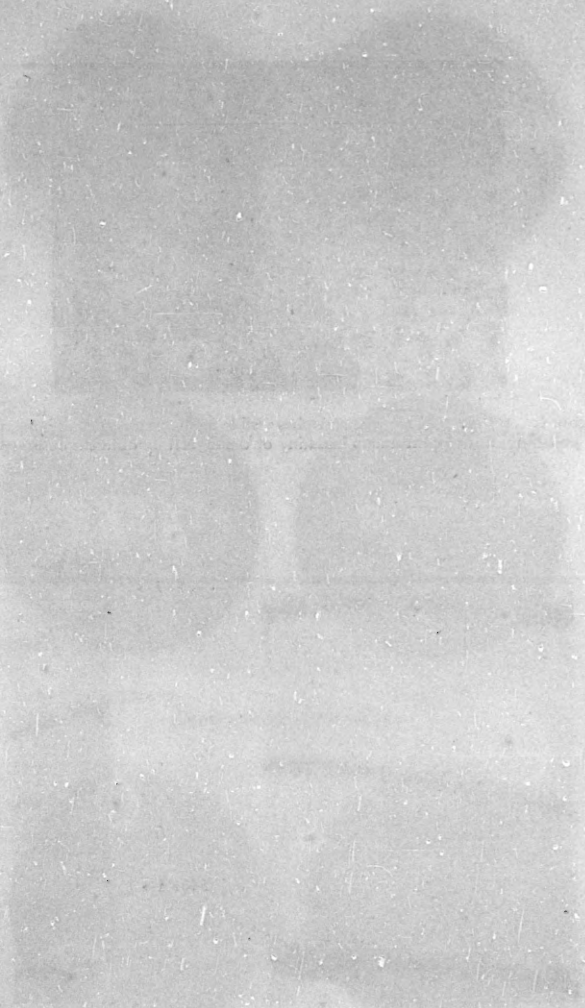


Figure 3. ×45.



Reactions between Oxide Cathodes and Gases at Very Low Pressures

By S. WAGENER*
Electronics Division, Post Office Research Station, Dollis Hill, N.W.2

MS. received 19*th November* 1953

Abstract. A method is described by which the behaviour of oxide cathodes under the impact of gases at pressures between 10^{-7} and 10^{-5} mm Hg can be studied. From the cathode under investigation saturated pulsed emission and space-charge-limited emission currents are drawn simultaneously. Whilst the pulsed emission is used as a measure of thermionic emissivity, the space-charge-limited emission is employed for operating an ionization gauge which indicates the gas pressure directly in front of the cathode. The experimental tube is connected to the pump via a capillary, and the rate of gas flowing into the cathode can be determined from measurements of the two pressures at either end of capillary.

The resistance of a cathode to attack by oxygen is found to be independent of the material used for the core but dependent to a considerable extent on the presence of small traces of other gases and of barium originating from the getter. The influence of emission current, cathode temperature and gas pressure on rate of poisoning by oxygen is examined.

By comparing the rate of oxygen flow into the cathode with the number of ions produced in the experimental tube, it is shown that the poisoning effect observed is not due to ions. From the rate of flow of oxygen the quantity of barium consumed during poisoning is calculated and found to be of the order of 0·01 mol %.

When exposure to oxygen ceases, complete recovery of the cathode is observed at temperatures between 1000 and 1100°K but not above 1200°K.

The effect of carbon dioxide on oxide cathodes is similar to that of oxygen but, at pressures below 5×10^{-6} mm, its poisoning effect is more pronounced for cathodes with passive cores. There are indications that carbon dioxide as opposed to oxygen reduces the emitting area of the cathode. Carbon monoxide poisons cathodes with active nickel cores but activates those with platinum cores.

Nitrogen does not affect the emission appreciably whilst water vapour, hydrogen and methane activate at pressures below 10^{-5} mm. The activating effect of methane is due to the reducing action of carbon which is deposited at the cathode surface owing to dissociation of the methane. The action of the carbon reverses into a poisoning one if the methane pressure exceeds the equilibrium pressure of the carbon monoxide originating from the reaction between carbon and barium oxide.

§ 1. Introduction

Among nearly 300 publications on oxide-coated cathodes which have been published during the last twenty years, only five are known to the author in which the effects of gases on such cathodes have been studied. It will be useful to examine the reasons for this disparity which contrasts with the

* Now with Kemet Company, Division of Union Carbide and Carbon Corporation, Cleveland, Ohio, U.S.A.

importance of gas effects for both performance and mechanism of oxide cathodes.

In practice the thermionic emission from oxide cathodes may vary considerably owing to chemical reactions occurring between the material of the cathode and residual gases. Since the pressure of these gases is of the order of 10^{-6} mm or lower (cf. Herrmann and Krieg 1941), investigation of such reactions should be undertaken in the same pressure range. If higher pressures are used some of the reactions occur so rapidly that details of the reaction process cannot be studied. Moreover, since the reactions concerned are based on a chemical equilibrium which is pressure dependent, they may at higher pressures take the direction opposite to that observed at lower pressures. Examples of this will be given below (§§ 6, 7).

If optimum experimental conditions are to be obtained, some further requirements must be satisfied. One of these arises from the considerable pressure differences which may exist in a vacuum system if movement of gases occurs at very low pressures. Owing to this the gas pressure must be measured in the immediate vicinity of the cathode and not at some place farther away. Furthermore, very small quantities of gas of the order of 10^{-6} cm^3 S.T.P. are required for producing the reactions in question. These quantities also have to be manipulated and measured.

Finally the strong temperature dependence of the reactions concerned does not justify extrapolation of results from one temperature range to another, unless the laws are found on which this temperature dependence is based. The temperature range which is of most interest is that in which oxide cathodes are normally operated, i.e. between 900 and 1100°K. Since steady saturated currents cannot be drawn from a fully activated cathode at these temperatures, it follows that pulsed emission currents should be used for measuring the thermionic emissivity of the cathode.

During recent years the author has developed a technique for manipulating and measuring very small quantities of gas at very low pressures (Wagener 1952, 1953). This technique which has previously been used for the investigation of getters, has been adapted to the study of reactions between oxide cathodes and gases. By combining this method with the measurement of pulsed emission currents drawn from the cathode under investigation, it has been possible to meet all requirements set out above.

§ 2. Experimental Details

The experimental tube (see figure 1 (*a*)) was made of a lower 'tetrode part', consisting of oxide cathode, two grids and anode, and an upper 'ionization gauge part' consisting of a collector for positive ions and a suppressor cylinder for suppression of undesirable photo-electrons from the collector (Metson 1951a). The ions used for the pressure measurement were produced by the anode current (normally $I_A = 5$ mA) in the interspace between screen grid and anode of the tetrode part, these two electrodes being connected to the same positive potential (normally 230 volts). The ionization gauge was calibrated with carbon monoxide, the calibration factor for this gas being found as $C^* = 0{\cdot}4$/mm Hg (cf. Wagener and Johnson 1951).

The experimental tube was connected to the manifold of a three-stage diffusion pump by a pumping tube (capillary) of 2 mm inside diameter and 40 mm length in the manner illustrated by Wagener (1952, figure 1). Gases could be

injected into the manifold via a needle valve. The rate R at which gas flowed into the oxide cathode could then be calculated from the flow resistance F of the pumping tube and from the two pressures p_{man} measured in the manifold, and p_t measured in the experimental tube:

$$R = F(p_{man} - p_t). \qquad \ldots\ldots(1)$$

The quantities of gas taken up by the cathode could be obtained by taking $\int R(t)\,dt$. (For further details see § 4.3.)

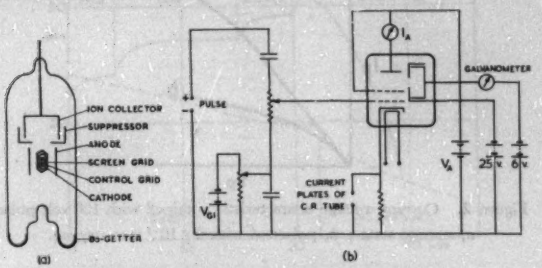


Figure 1. Principle of experimental arrangement.
(*a*) experimental tube; (*b*) circuit diagram.

The circuit diagram of figure 1(*b*) shows how pressure measurement and measurement of thermionic emissivity were combined. The negative d.c. potential which was applied to the control grid, for adjustment of the d.c. anode current to its desired value, was superimposed by a positive, saw-tooth shaped voltage pulse of 2 μsec length and 50 sec^{-1} pulse repetition frequency. During the duration of this pulse the cathode emits an emission current, the peak value of which will be saturated provided that the peak voltage has been chosen high enough. This peak current was taken as a measure of the thermionic emission from the cathode. It was realized that this current could not readily be converted into work functions or other emission constants, but this value was preferred for the present investigation because it could be measured in a simple manner by means of a valve voltmeter.

The pulsed voltage was produced by a gas-filled trigatron which in spite of its lack of stability was adequate for these first experiments. The peak voltages employed were either 100 or 250 volts, depending on temperature of the cathode and magnitude of its emission. As an example figure 2 shows two current–voltage characteristics displayed on a cathode-ray tube, one for the normal state of the cathode and one for exposure to oxygen.

The range of cathode temperatures which could be used was limited at the upper end by the temperature at which saturation could no longer be obtained with 250 volts. The application of still higher pulsed voltages was rejected because it was felt that pulsed voltages should not appreciably exceed the d.c. voltages. At the lower end the temperature limit was reached when the emission currents became too small to be measured with the valve voltmeter without increasing the resistor in the cathode lead beyond the impedance value of the stray capacitance. The temperature range obtained in this way was between

1000 and 1075°K. The lower limit, of course, could be extended by using a higher amplification for the valve voltmeter.

The oxide cathodes employed had a coating made by spraying barium and strontium carbonates in equimolar proportions (0·5 cm² area, 3·5 mg weight, 100 μ thickness). The cores used for supporting this coating consisted either of active nickel, containing 0·1% magnesium and 0·05% silicon, or of platinum.

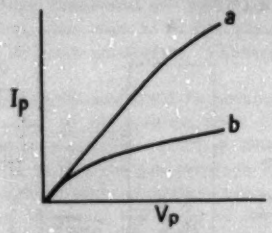


Figure 2. Current–voltage characteristic obtained with 100 volt pulse.
a, normal state ; *b*, poisoned with 8×10^{-7} mm oxygen.

The experimental tubes, after being sealed to the pump, were submitted to the usual processes, baking, radio-frequency heating of metal parts, decomposition of the carbonates of the cathode coating into oxides and bombardment of metal parts by electrons. In some cases a barium getter was flashed at the end of the pumping process (see figure 1 (*a*)). This had the advantage that lower pressures could be obtained in the tube, but the disadvantage that, because of the additional flow of gas into the getter, the rate of flow into the cathode could no longer be calculated separately.

The gases to be investigated were selected according to the likelihood of their occurrence in tubes using oxide cathodes. Oxygen was either made from potassium permanganate or taken from steel cylinders, no difference in behaviour being found. Carbon dioxide was prepared either from magnesium carbonate or calcium carbonate, and carbon monoxide from a mixture of 5·2 parts zinc powder and 4 parts calcium carbonate (Weinhouse 1948). Hydrogen was supplied through a palladium tube, and water vapour was obtained from distilled water which was frozen in liquid air several times while the air above it was pumped away. Nitrogen and methane were obtained from steel cylinders. All gases except H_2O were passed over phosphor pentoxide previous to use.

§ 3. General Phenomena Observed

3.1. *Transient Phenomena*

During the first investigations barium getters were used in the experimental tube. When exposing the cathodes in such tubes to oxygen it was found that an unexpectedly high oxygen pressure of nearly 10^{-5} mm was required in order to obtain any reduction in the emission current I_p (see figure 3). At lower pressures an increase in emission was observed.

It was soon recognized that this behaviour was somehow caused by the flashing of the getter. If, after the first exposure to oxygen, the emission was allowed to recover and the cathode subsequently exposed to oxygen again, an

entirely different behaviour was observed (see dashed curve of figure 3), the emission now beginning to fall at pressures of a few times 10^{-7} mm. Obviously, during flashing of the getter, a certain small quantity of barium, protecting the cathode from attack by oxygen, had been deposited on the cathode surface. It is interesting to note that this protective barium did not increase the emission from the cathode which, after recovery from the first exposure to oxygen, came up to the same value as before.

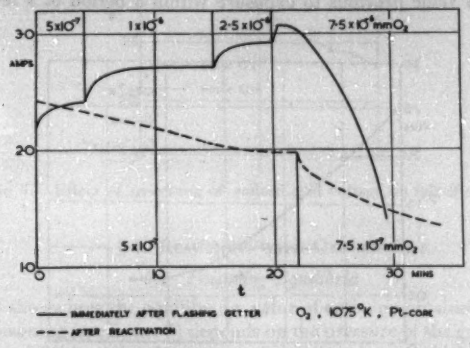


Figure 3. Effect of flashing a getter on sensitivity to exposure to oxygen.

Similar effects were observed with other gases. After they had been recognized, the getters were either flashed very slowly or, better still, the cathode was first exposed to a certain dose of oxygen, removing the protective surface barium before the proper measurements were commenced.

Even after avoiding this effect produced by the getter, a slight increase in emission was frequently observed when the cathode was exposed to oxygen at very low pressures (a few times 10^{-7} mm). This increase, which is illustrated in figure 4, was not permanent. After having passed a peak value the emission

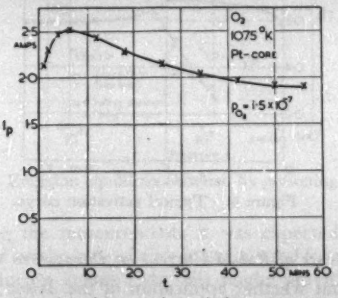


Figure 4. Transient increase in emission produced by oxygen.

fell again and finally reached an equilibrium value which was lower than the initial one. This activation, produced by very low oxygen pressures, is obviously the same as that described by Liebold (see Herrmann and Wagener 1951, p. 271).

3.2. *The Two Types of Reaction Curves*

After the transient phenomena described above had been excluded, all phenomena observed were reproducible. The gases could then be classified into two groups—poisoning gases which reduced the emission and activating ones which increased it. A typical poisoning curve is shown in figure 5. After a certain period the emission reaches an equilibrium value at which it stays as long as the pressure is not altered. If the gas is pumped off completely the emission recovers to its value previous to exposure within a period of a few hours (see figure 9).

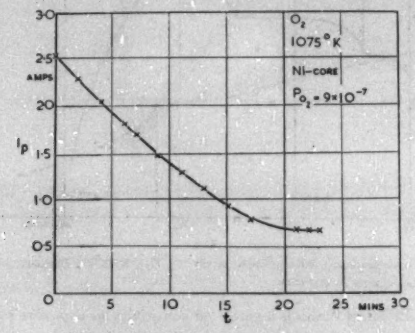


Figure 5. Typical poisoning curve.

If activation occurs, the behaviour of the cathode, as shown by figure 6, is exactly the opposite.

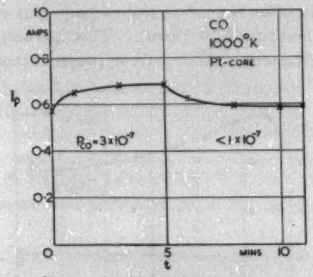


Figure 6. Typical activation curve.

3.3. *Influence of Pulsed Current on Phenomena Observed*

In order to find out whether application of the 2 μsec pulse influences the phenomena observed, the pulsed voltage was switched off for intervals of 2 minutes' duration. Figure 7 shows a poisoning curve obtained with oxygen in this way. It is seen that the rate of poisoning does not change appreciably during

the off-periods. In the same manner it was found that the equilibrium values measured after a certain period of exposure to gas are not affected by application of the pulse.

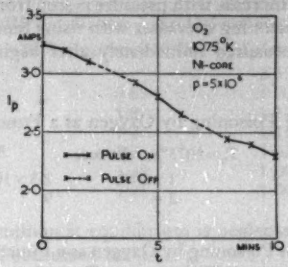


Figure 7. Effect of switching off pulsed grid voltage on fall of emission.

§ 4. Reactions with Oxygen

4.1. *Poisoning Equilibria*

Figure 8 shows how the equilibrium value of emission obtained after a certain period of poisoning (see figure 5) depends on the pressure of the gas. The values are well reproducible and independent of the direction in which the pressure is varied. A variation of temperature within the range available for the measurements (1000–1075°K) did not alter the equilibrium values appreciably.

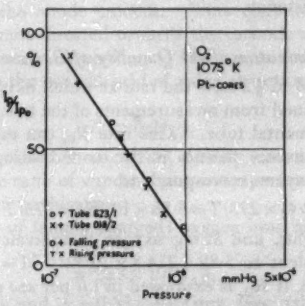


Figure 8. Emission equilibria obtained by poisoning with oxygen.

Before beginning the measurements it was expected that the equilibrium values would depend on the type of core material used for the cathode. A nickel core containing reducing agents (Mg, Si) might provide the cathode with a higher resistivity to poisoning by oxygen. However, no appreciable difference was found between the two core materials when using oxygen. Possibly this result is due to the fact that nickel was used for other components of the experimental tube from which reducing material may have migrated to the cathode.

4.2. *Rates of Poisoning*

The rate at which poisoning occurs was investigated as a function of three parameters, oxygen pressure, temperature of the cathode and d.c. current drawn from the cathode. The increase with pressure is seen from table 1, whilst table 2 shows that the rate of poisoning decreases with rising temperature. The values given are initial ones measured immediately after beginning the exposure to oxygen.

Table 1. Rate of Poisoning by Oxygen as a Function of Pressure

T_c=1075°K, Pt-core

Pressure (mm Hg)	1×10^{-4}	$2{\cdot}5\times10^{-4}$	6×10^{-4}
$(\Delta I_p/I_p)/\Delta t$ (%/min)	3·4	9	14

Table 2. Rate of Poisoning by Oxygen as a Function of Cathode Temperature T_c

T_c (°K)		990	1025	1055	1085
$\frac{\Delta I_p/I_p}{\Delta t}$ (%/min)	Cathode 1	11·2	8·8	6·4	—
	Cathode 2	—	8·5	—	3·2
$\frac{T_c}{I_p}\times\frac{\Delta I}{\Delta t}\times10^{-3}$	Cathode 1	11·1	9·0	6·7	—
	Cathode 2	—	8·7	—	3·5

Consistent results could not be obtained for the dependence on d.c. emission current. In some cases the rate of poisoning decreased when the d.c. current (anode current) was increased; in others exactly the opposite behaviour was observed. More detailed investigations will be necessary in order to find the reasons for this inconsistency.

4.3. *Determination of the Quantity of Barium Consumed*

It has been outlined in §2 how the rate at which oxygen is taken up by the cathode can be determined from measurements of the two pressures in the manifold and in the experimental tube. This rate R_{O_2} can easily be converted into the rate R_{Ba} at which excess barium in the oxide coating is consumed by the oxygen. Since 1 l.μ oxygen corresponds to

$$M/(760\times22{\cdot}4)\times273/T=5{\cdot}88\times10^{-5}M\times273/T \text{ mg oxygen}$$

(M the molecular weight), and M mg oxygen is equivalent to $2\times137{\cdot}4$ mg Ba, 1 l.μ oxygen is equivalent to $5{\cdot}88\times274{\cdot}8\times10^{-5}\times0{\cdot}91=1{\cdot}47\times10^{-2}$ mg Ba (for T=300°K). Therefore, if R_{O_2} is expressed in l.μ per sec and R_{Ba} in μg per sec

$$R_{Ba}=14{\cdot}7R_{O_2}. \qquad \ldots\ldots(2)$$

Table 3 shows an example of a determination of R_{Ba}. As the poisoning equilibrium is approached, the values calculated for R_{Ba} decrease. This confirms that the oxygen flowing through the capillary into the experimental tube indeed combines with the active barium responsible for the emission. The fact that the values obtained for R_{Ba} are not equal to zero when the equilibrium is reached, may be due either to an additional adsorption of oxygen by other parts of the tube or to insufficient accuracy of the method. The pressure difference from which the rate of consumption is calculated is comparatively small. By reducing the

diameter of the pumping tube this pressure difference could be increased and the consumption rate at the equilibrium be determined more accurately.

Table 3. Rate of Consumption of Barium during Poisoning

$T_c = 1075°K$, Ni-core $F = 17{\cdot}5\ cm^3\ sec^{-1}$

Time (min)	p_{man} (mm × 10^{-6})	p_t (mm × 10^{-6})	I_p (amps)	R_{Ba} (10^{-4} μg sec^{-1})
—			2·45	
½	3·2	0·85	1·90	6·0
2	3·2	1·0	1·48	5·6
4	3·3	1·2	1·15	5·4
7	2·55	1·3	1·02	3·2
10	2·1	1·35	1·00	1·9
14	2·0	1·25	1·00	1·9

If the rate of consumption at equilibrium is neglected, a rough value of the total quantity of barium consumed is:

$$Q_{Ba} \simeq \tfrac{1}{2}(6{\cdot}0 - 1{\cdot}9) \times 10^{-4} \times 600 = 0{\cdot}12\,\mu g.$$

In another case where the emission was reduced to 0·3 amp a value $Q_{Ba} = 0{\cdot}5\,\mu g$ was obtained. This, with 3·5 mg weight of the coating as carbonates, corresponds to about 2×10^{-2} mol %. This value is of the same order as the total quantity of excess barium in an oxide cathode which was determined by Jenkins and Newton (1949) by a volumetric method.

4.4. *Rates of Poisoning: Theoretical Aspects*

Based on the experimental data given in the previous subsections, a few theoretical aspects may be discussed. One of these is whether the gas particles poisoning the cathode are molecules or ions produced by collision with the electrons constituting the anode current. This question can be decided by comparing the ion current directed towards the cathode with the rate at which excess barium is consumed. According to Herrmann and Krieg (1941) the ion current I^+ flowing into the control grid of valve structures similar to that used here, is given by

$$I_{G1}^+ \simeq 0{\cdot}2 I_A p_t \qquad \ldots\ldots(3)$$

(p_t in mm Hg). From this the ion current flowing into the cathode is obtained by multiplying with the ratio of winding wire interspace to wire diameter of the control grid which is 2·5 in this case.

Hence, the number of (singly charged) oxygen ions hitting the cathode per second is

$$n^+ \simeq 3 \times 10^{18} I_A p_t, \qquad \ldots\ldots(4)$$

or with $I_A = 0{\cdot}005$ amp and $p_t = 0{\cdot}85 \times 10^{-6}$ mm (see table 3), $n^+ = 1{\cdot}3 \times 10^{10}$ ions sec^{-1}. This value corresponds to

$$1{\cdot}3 \times 10^{10} \times 2 \times 137{\cdot}4 \times 1{\cdot}66 \times 10^{-18} \simeq 0{\cdot}06 \times 10^{-4}\,\mu g\,sec^{-1} \text{ barium.}$$

Comparing with the first line of table 3, it is seen that only 1% of the barium consumption measured can be due to ions. Poisoning by oxygen, therefore, is produced by neutral particles.

The rate at which molecules hit the cathode surface is given by the well known equation:

$$n = 3{\cdot}51 \times 10^{22} p/(MT)^{1/2} \text{ molecules } cm^{-2}\,sec^{-1} \qquad \ldots\ldots(5)$$

(cf. Dushman 1949, p. 17). With $M=32$, $T=300°\text{K}$, $p=0{\cdot}85\times10^{-6}$ mm and for 0·5 cm² area, we obtain $n=1{\cdot}5\times10^{14}$ molecules sec^{-1} corresponding to 700 μg sec^{-1} barium. It follows therefore that only 1% of the molecules impinging upon the surface reacts with the excess barium. This rather small percentage may be explained in two different ways. Either only 1% of the surface area of the cathode is liable to be affected by impinging oxygen molecules, or the particles which produce poisoning are not molecules but atoms which are created by collisions between a small percentage of the gas molecules and electrons constituting the anode current. The second assumption is not very likely to be true, because no fundamental difference has been found between poisoning phenomena measured with larger voltages and those observed with very small voltages which are insufficient for creating atoms from molecules.

Another aspect to be discussed is whether poisoning by oxygen is only a surface phenomenon or whether processes in the interior of the oxide coating are an essential part of it. This question can be decided by considering the temperature dependence of the rate of poisoning.

If poisoning were only a surface phenomenon in which barium atoms or ions at the surface combine with oxygen molecules arriving from the gas phase, the rate at which barium is consumed at the beginning of poisoning would only depend on the rate of arrival of molecules. Since this rate of arrival is independent of the temperature of the cathode, one would expect the initial rate of barium consumption to be temperature independent as well. The same conclusion would then be valid for the rate at which the work function of the cathode increases, because this work function depends on the quantity of excess barium. Finally, the correlation between rate of increase of work function and rate of fall of emission current is obtained by differentiating the emission equation:

$$I_p=AT_c^n\exp(-b/T_c), \qquad \ldots\ldots(6)$$

in which b is a measure of the work function. Differentiation gives

$$\frac{db}{dt}=-\frac{T_c}{I_p}\frac{dI_p}{dt}. \qquad \ldots\ldots(7)$$

According to this, if poisoning were a mere surface phenomenon, values of $(T_c/I_p)\times(\Delta I/\Delta t)$ should not depend on temperature. Table 2, however, shows that these values fall with increasing temperature to a considerable extent. This fall can be explained by assuming that with increasing temperature either a greater part of the oxygen arriving at the surface diffuses into the interior where it does less damage to the emission, or that a greater part of the barium at the surface which is consumed by oxygen is replenished by barium diffusing from the interior towards the surface. Measurements of barium consumption as a function of temperature should give more evidence on these processes (also see § 8).

4.5. *Reversible and Irreversible Poisoning*

The recovery from oxygen poisoning has been investigated in several publications (Herrmann and Krieg 1949, Metson 1949, 1951b, Shepherd 1953). All these investigations were undertaken without application of ionizing voltages. In the experiments described here recovery was studied with all voltages (anode voltage, pulsed voltage) connected. Under these conditions, using maximum oxygen pressures of about 10^{-5} mm, complete recovery was always observed, even after several exposures to oxygen. Figure 9 shows an example in which the

fall in emission during poisoning extended from 3·5 to 0·02 amp. In this case the final emission current measured after recovery even exceeded the initial current observed before poisoning. Recovery occurred independently of the type of core material used for the cathode.

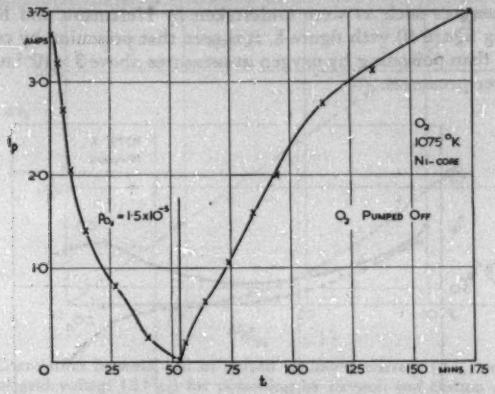


Figure 9. Poisoning and recovery of emission obtained by exposure to oxygen.

In § 4.3 it was shown that the quantity of excess barium, consumed when the cathode is poisoned to about one tenth of its initial emission, is of the same order of magnitude as the total quantity of barium contained in the cathode. Since, on the other hand, recovery occurs after quite a number of such poisonings, the recovery phenomenon cannot be explained by the existence of a barium reservoir from which the barium, consumed during previous poisoning, is replenished. It could be assumed that recovery is due to electrolysis of the oxide coating but this assumption has been ruled out by Metson (1949) who showed that recovery can be obtained without drawing any current from the cathode. Hence, it appears that oxygen does not combine permanently with the excess barium by forming barium oxide, but goes into a less permanent association from which the oxygen particles are easily released when the pressure is reduced. As suggested by Herrmann and Wagener (1951, pp. 268–70) such a loose association may be due to the presence of singly charged O^--ions which are more easily formed from impinging oxygen molecules than are the doubly charged O^{2-}-ions necessary for formation of barium oxide. The release of O^--ions during recovery has been confirmed experimentally by Shepherd (1953) by means of a mass spectrometer.

It should be emphasized that the complete reversibility of oxygen poisoning such as described above is only observed in a certain temperature range (e.g. 1000–1075°K). If the cathode is poisoned at, for instance, 1250°K reversibility no longer exists. In such a case the emission value, measured after 30 minutes poisoning in 2×10^{-5} mm oxygen and a subsequent recovery time of several hours, was only one tenth or less of the initial value. There were no signs of any further recovery. At these higher temperatures, therefore, the association between oxygen and excess barium appears to be permanent. A tentative explanation for this behaviour will be given in § 8.

§ 5. Reactions with Carbon Oxides

The effect of carbon oxides on the emission from cathodes with nickel cores is seen from figure 10. The curves show clearly that the reduction in emission measured between 10^{-7} and 10^{-5} mm cannot be predicted from measurements at higher pressures such as were undertaken by Herrmann and Krieg (1949). By comparing figure 10 with figure 8, it is seen that poisoning by carbon oxides is less severe than poisoning by oxygen at pressures above 3×10^{-7} mm but more severe at lower pressures.

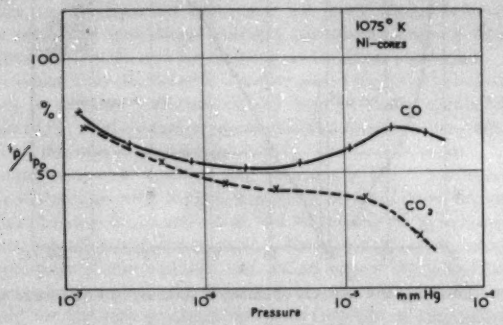


Figure 10. Emission equilibria obtained by poisoning with carbon oxides.

With carbon oxides a definite influence of core material on the reaction between oxide cathode and gas was observed. Carbon dioxide produced a much stronger effect on cathodes with platinum cores than on those with active nickel cores. During exposure to 3×10^{-7} mm carbon dioxide for instance, the emission from cathodes with nickel cores fell by about 40% while that from cathodes with platinum cores fell by 90%.

There is one further aspect of the poisoning effect produced by carbon dioxide which is of some interest. When exposing cathodes to oxygen, the reduction in pulsed emission was not accompanied by a corresponding reduction of the d.c. anode current during the initial phase of poisoning. This behaviour is such as will be expected. The explanation for it is that the anode current is space-charge-limited and that, contrary to a saturated pulsed current, a space-charge current is very little influenced by variations in work function of the cathode, unless the order of magnitude of the saturated current becomes the same as that of the space-charge current.

During exposure to carbon dioxide a very different behaviour was observed. The anode current fell right from the beginning of exposure, and this fall had to be compensated by an adjustment of the d.c. bias. These observations are illustrated by figure 11 in which the variation in bias (ΔV_{G1}) necessary for maintaining constant anode current has been plotted against the percentage fall of pulsed emission.

The behaviour observed with carbon dioxide can be explained if it is assumed that the cathode is not homogeneously affected by this gas, but that only certain preferred parts of the emitting cathode area are poisoned while others go on

emitting in the same way as before. Such a loss of emitting area would result in a simultaneous fall of pulsed emission and space-charge emission as observed.

The results obtained when studying the impact of carbon monoxide on cathodes with platinum core were very surprising. Instead of a poisoning

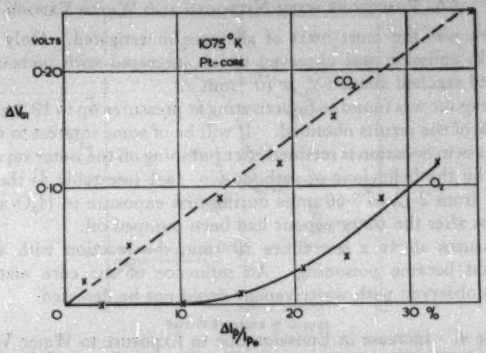


Figure 11. Correlation between fall in pulsed emission current (ΔI_p) and variation of control grid voltage (ΔV_{g1}) for poisoning by oxygen and carbon dioxide.

effect as with nickel core cathodes, a very marked activation was observed. In figure 12 such activation is illustrated and compared with the poisoning observed for a nickel core cathode. At first this result was not believed and was attributed to fortuitous experimental conditions. Any doubts, however, had to disappear in the end because this difference in reaction to carbon monoxide could be confirmed by investigation of five cathodes each of which were made with the two core

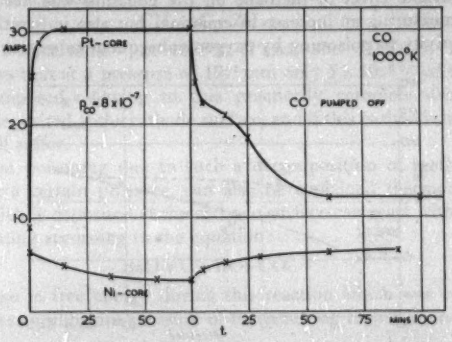


Figure 12. Effect of carbon monoxide on cathodes with different cores.

materials, and could even be used for analysing the core material in cases where its composition was not known.

The results described above suggest that the core material may have a more direct influence on the electron emission from the cathode than is normally assumed.

Any explanation of the phenomena observed with carbon oxides will have to be proved by a separate and detailed investigation. No attempt, therefore, will be made here to offer such an explanation.

§ 6. Reactions with Nitrogen and Water Vapour

Nitrogen was the most inert of all gases investigated. Only a very small reduction in emission was observed which increased with increasing nitrogen pressure and reached about 5% at 10^{-5} mm.

Water vapour was found to be activating at pressures up to 10^{-5} mm. Table 4 shows a few of the results obtained. It will be of some interest to note that part of the increase in emission is retained after pumping off the water vapour. This is illustrated by the behaviour of cathode No. 18/1 (see table 4) the emission of which rose from 2·12 to 3·60 amps during first exposure to H_2O and remained at 3·30 amps after the water vapour had been pumped off.

At pressures above a few times 10^{-5} mm the reaction with water vapour reversed and became poisoning. An influence of the core material on the phenomena observed with water vapour could not be detected.

Table 4. Increase in Emission due to Exposure to Water Vapour

$p_{H_2O} = 5 \times 10^{-6}$ mm, $T_c = 1075°K$

Cathode No.		16/2	18/1	18/1	18/4
Emission (amps)	before exposure	3·42	2·12	3·30	0·50
	during exposure	3·50	3·80	3·60	3·25

§ 7. Reactions with Hydrogen and Methane

Both hydrogen and methane were found to be activating, the results obtained with one particular cathode being shown in figure 13.

The favourable effect of methane on the emission was not only observed directly, by measuring an increase in emission, but also indirectly by detecting a higher resistivity to poisoning by oxygen subsequent to exposure to methane.

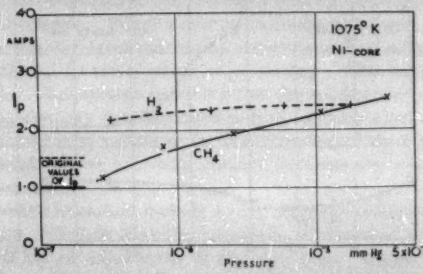


Figure 13. Increase in emission produced by hydrogen and methane.

This behaviour is illustrated in figure 14. After exposing the cathode concerned to 10^{-5} mm methane, the methane was pumped off until a residual pressure of 5×10^{-7} mm was obtained. Subsequent exposure to oxygen gave poisoning equilibria which were far to the right of those obtained normally (cf. figure 8).

After pumping off the oxygen, the procedure was repeated twice, every time beginning with a lower residual pressure (see figure 14). Gradually the equilibrium emission values due to poisoning moved towards lower pressures and approached those measured without previous exposure to methane.

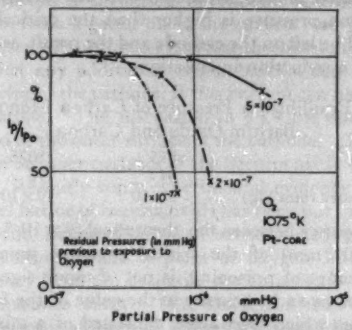


Figure 14. Influence of previous exposure to methane on sensitivity to oxygen.

Contrary to that of hydrogen, the activating effect of methane was only observed at low pressures up to a limit of about 5×10^{-5} mm. Above this value methane became poisoning, the poisoning effect increasing with increasing pressure. This poisoning effect which was also observed by Herrmann and Krieg (1949) can be explained by the fact that methane decomposes readily into carbon and hydrogen at temperatures of the order of 1000°K (Rittner 1953). The degree to which such a decomposition occurs can be calculated from the change in free energy during the reaction

$$CH_4 = C + 2H_2; \qquad \ldots\ldots(8)$$

this is $\Delta F = -5{\cdot}3$ kcal mole^{-1} at 1075°K (Parks and Huffman 1932). The calculation shows that at a pressure of 10^{-4} mm only 5×10^{-6}% of the methane remain undecomposed. Owing to this practically complete decomposition, carbon will be deposited at the cathode surface, and if this carbon is not removed, the emission will suffer.

The fact that poisoning due to such a decomposition of methane is only observed above a certain pressure, can also be explained thermodynamically. The carbon which is deposited at the cathode surface can react with the barium oxide of the coating according to the equation:

$$BaO + C = Ba + CO. \qquad \ldots\ldots(9)$$

From the change in free energy during this reaction which was computed by White (1949) the equilibrium pressure of the resulting carbon monoxide can be calculated:

$$\log p_{CO} = -\frac{\Delta F}{9{\cdot}14\,T_c/1000} + 2{\cdot}88. \qquad \ldots\ldots(10)$$

The pressures obtained are given in table 5.

When the cathode is exposed to methane of a certain pressure, the quantity of carbon deposited by the methane is stoichiometrically equivalent to a quantity

of carbon monoxide of the same pressure as the methane. If, therefore, the methane pressure is lower than the equilibrium pressure of carbon monoxide (9×10^{-6} mm at 1075°K), all the carbon deposited will immediately react with barium oxide without the equilibrium pressure of carbon monoxide being reached. The result is formation of excess barium and activation of the cathode. If, however, the methane pressure is higher than the critical value, some of the carbon deposited will be left on the cathode and the result, as shown by Herrmann and Krieg (1949), is a reduction in emission.

Table 5. Equilibrium Pressure of Carbon Monoxide above Barium Oxide and Carbon

Temperature (°K)	875	1075	1275
ΔF (kcal mole^{-1})	91	78	65
Equilibrium pressure (mm Hg)	3×10^{-9}	9×10^{-6}	2×10^{-3}

The slight discrepancy between the theoretical (9×10^{-6} mm) and the experimental value (5×10^{-5} mm) of the critical pressure separating the range of activation from the range of poisoning, is not of much significance. It may be explained for instance by an inaccuracy in the value of the free energy, or by the fact that the cathodes which were used consisted of a mixture of barium and strontium oxide instead of barium oxide only.

The existence of a critical pressure for activation by methane explains why some of the previous attempts to use methane activation have not been very successful and only produced rather low emission values.

In concluding this section some common features should be mentioned which are observed when activating an oxide cathode by exposure to a gas. In all cases investigated the behaviour of the emission current was such as illustrated by figures 6 and 12. During exposure the emission rose to a maximum value, while afterwards, when the gas was pumped off, the emission fell to a value which was lower than the maximum but still higher by a certain amount than the initial value measured before exposure. Obviously the quantity of excess barium produced in the presence of the gas cannot be maintained when the gas is absent, and is removed, probably due to evaporation. It should be emphasized that it is the maximum values measured in the presence of the gas which have been plotted in figure 13.

As for the final emission measured after exposure to the activating gas, it was never possible to exceed a certain limiting value which with the majority of the cathodes could be reached by normal activation methods without using a gas. Under the existing conditions this limit was about 3 amps at 1075°K. Owing to this limitation activation by means of gases was most successful, if, due to certain difficulties during processing of the tube, the initial emission was rather low. (See last column of table 4.)

§ 8. Conclusions

Among the gases investigated there are only three, oxygen and the two carbon oxides, which reduce the emission of oxide coated cathodes in the pressure range below 10^{-5} mm*. The pressure limit below which the deleterious effect ceases

* To these detrimental gases, chlorine must be added; this reduced the emission current to 35% at a pressure of 4×10^{-7} mm and to 10% at 10^{-6} mm. Sulphur dioxide, however, did not affect the emission appreciably at pressures below 10^{-5} mm.

could only be measured for oxygen; it is 10^{-7} mm. For carbon oxides this limit may well be one or two decades lower.

The experiments indicate that the quantity of excess barium in or on an oxide cathode is not the only factor determining its thermionic emission. If the quantity of excess barium is increased beyond an optimum value, either by evaporation of barium from a getter on to the surface (see § 3.1) or by exposure to an activating gas (see § 7), a further increase in emission is not obtained. It may be concluded that any additional excess barium can only be utilized for increasing the emissivity of the cathode, if this barium can be deposited in special sites of the crystal lattice. The discussion in § 4.4 has shown that these sites cannot only be located at the outer surface of the cathode. A major part of them must be located either at inner surfaces of the barium oxide crystals or inside the crystals themselves. Recently some experimental evidence for the existence of vacancies in the crystal lattice of barium oxide has been put forward by Redington (1952) and Sproull and Bever (1953) and substantiated by Nergaard (1952). The special sites, therefore, could be pairs of oxygen- and barium vacancies. As long as such vacancy pairs are available the emission could be increased by building barium atoms into them. The conclusion drawn in § 4.4 that only 1% of the cathode surface is affected by oxygen then would mean that the surface density of such vacancies is 1%.

The diffusion measurements by Redington (1952) and Bever (1953) have shown that the number of vacancies in barium oxide or barium–strontium oxide is structure-sensitive at temperatures below about 1300°K. At lower temperatures, therefore, this number is determined by the processes defining the structure of the barium oxide, for instance the process of precipitating the carbonate crystals and the process of decomposing these into oxides. Hence, with given processing methods, the number of available vacancy pairs will always be about the same, and the maximum emission obtainable by any activation method will also be the same. This is what was actually observed.

The number of vacancy pairs may be influenced if the temperature of the cathode is raised beyond the structure-sensitive range to temperatures at which vacancies which are due to structure irregularities can be removed. Such a removal of vacancies resulting in an irreversible loss of emission was indeed observed after exposing cathodes to oxygen at higher temperatures (see § 4.5).

Acknowledgments

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in this paper. The author wishes to thank Messrs. R. E. Hines and R. E. Thorne for their skilled assistance in the experimental work.

References

(*a*) *Referring to reactions between oxide cathodes and gases.*

Arizumi, T., and Narita, S., 1948, *J. Phys. Soc. Japan*, **3**, 356.

Hamaker, H. C., Bruining, H., and Aten, A. H. W., 1947, *Philips Res. Rep.*, **2**, 171.

Herrmann, G., and Krieg, O., 1949, *Ann. Phys., Lpz.*, **4**, 441.

Metson, G. H., 1949, *Nature, Lond.*, **164**, 540.

Shepherd, A. A., 1953, *Brit. J. Appl. Phys.*, **4**, 70.

(*b*) *Others.*

BEVER, R. S., 1953, *Brit. J. Appl. Phys.*, **24**, 1008.

DUSHMAN, S., 1949, *Scientific Foundations of Vacuum Technique* (London : Chapman and Hall ; New York : John Wiley).

HERRMANN, G., and KRIEG, O., 1941, *Telefunkenröhre*, **21/22**, 219.

HERRMANN, G., and WAGENER, S., 1951, *The Oxide-Coated Cathode* (London : Chapman and Hall).

JENKINS, R. O., and NEWTON, R. H. C., 1949, *J. Sci. Instrum.*, **26**, 172.

METSON, G. H., 1951 a, *Brit. J. Appl. Phys.*, **2**, 46 ; 1951 b, *Vacuum*, **1**, 283.

NERGAARD, L. S., 1952, *R. C. A. Review*, **13**, 464.

PARKS, G. S., and HUFFMAN, H. M., 1932, *The Free Energies of Some Organic Compounds* (New York : The Chemical Catalog Co.).

REDINGTON, R. W., 1952, *Phys. Rev.*, **87**, 1066.

RITTNER, E. S., 1953, *Philips Res. Rep.*, **8**, 184.

SPROULL, R. L., and BEVER, R. S., 1953, *Phys. Rev.*, **90**, 352.

WAGENER, S., and JOHNSON, C. B., 1951, *J. Sci. Instrum.*, **28**, 278.

WAGENER S., 1952, *Proc. Instn Elect. Engrs*, Pt. III, **99**, 135 ; 1953, *Proc. Phys. Soc.* B, **66**, 400.

WEINHOUSE, T., 1948, *J. Amer. Chem. Soc.*, **70**, 442.

WHITE, A. H., 1949, *Brit. J. Appl. Phys.*, **20**, 856.

Some Measurements of the Electrical Conductivity, Thermoelectric Power and other Properties of the Coating in the Oxide-Coated Cathode

By C. M. LOVETT
Post Office Engineering Dept., Research Station, Dollis Hill, London N.W.2

M.S. received 12*th October* 1953, *and in amended form* 11*th January* 1954

Abstract. Measurements of the electrical conductivity of barium–strontium oxide made by a number of investigators have led to different observations as to its temperature dependence, from which, in some cases, it has been concluded that two characteristic work functions exist and therefore a complex mechanism of conduction. In this investigation measurements of the conductivity have been carried out under carefully controlled conditions and the conclusion reached that, for the range of temperature considered, only one work function exists. This was confirmed by the thermoelectric power measurements.

From comparison between the internal coating conductance and the anode–cathode conductance it is concluded that low electron emission is associated with a high resistive surface layer of high work function.

§ 1. Introduction

Various investigations into the electrical properties of the oxide coating of the oxide-coated cathode have been carried out recently which, using different methods, have led to different results. Those of Loojes and Vink (1949), and Young (1952), in which a block of the oxide between two nickel plates was used, led to the conclusion that two internal work functions exist, one referring to temperatures less than 800°K and the other to those greater than 800°K. In the measurements of Hannay, McNair and White (1949) the proportionality between conductivity and electron emission current was established using a specimen of barium–strontium oxide on a magnesia tube. The emission current was measured in a direction perpendicular to that of the conductivity measurements; for temperatures less than 1100°K only a single value of the work function was found. Metson, Wagener, Holmes and Child (1952) deduced the cathode resistance from the deviation of the experimentally determined retarding field characteristics from the theoretical value. As commercial pentodes were used in these experiments the double value of the work function observed was attributed to the 'interface' resistance.

In the present investigations a comparison has been made between the internal work function, as determined from the conductance and thermoelectric power, and the total work function as determined from the total emission current, and between the internal conductance and the total anode–cathode conductance, all these quantities being measured in the same direction.

Pure platinum was used as a core material in order to minimize the possibility of reducing agents being present. These are known to increase the tendency of the formation of a high resistive interface layer on the core.

§ 2. Tube Design

The tube consists essentially of a diode with probes embedded in the barium–strontium oxide coating of the cathode.

A thermo-pure platinum cylinder P (figure 1 (*a*)), of diameter 3·0 mm, length 30 mm and wall thickness 0·05 mm, is fixed as a push-fit on an alumina tube R and the two mounted by means of nickel wires on the support consisting of the chrome–iron wires I and glass insulating beads G. Two molybdenum wire springs S_1 and S_2, the ends of which are insulated from each other by means of the glass beads G′ and carry four small masts M_1, M_2, M_3, M_4, are mounted at right angles to the main support, so enabling a thin tape 'probe' T to be fixed under tension round the cylinder P. A 13% rhodium–platinum tape 12 μ thick and 0·165 mm wide is welded to the point A (figure 1(*b*)), so making

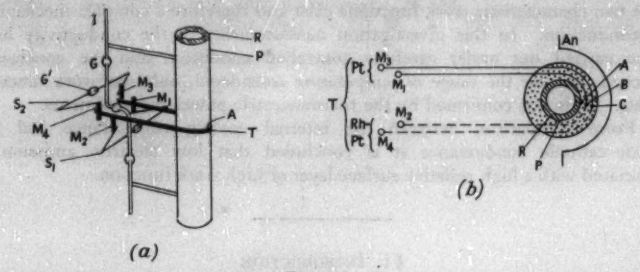


Figure 1.

a thermojunction at this point. A length of 5 mm of the platinum cylinder about A is then coated to a thickness of 40 μ with barium–strontium carbonate. The 'probe' B, consisting of a thermo-pure platinum and a 13% rhodium–platinum tape, is then welded to the masts M_1 and M_2, with the spring S_1 displaced to give a tension of 30 grammes on release, this value being determined by the tensile strength of the platinum at the temperatures to which it will subsequently be submitted. A second layer of barium–strontium carbonate is then sprayed on and the tape 'probe' C welded to the masts M_3 and M_4; finally a coating 20 μ thick is sprayed over the whole. The cathode so formed is mounted on a special valve pinch carrying two thermo-pure platinum and two 13% rhodium–platinum wires. A 4-watt valve heater is inserted in the hollow alumina cylinder R. The grid-type anode An, consisting of five molybdenum wires bent in the shape of a semicircle as shown, is mounted approximately 1·0 mm from the cathode surface. The diode electrode assembly so formed is then sealed into a valve envelope with two small branch tubes at the top of its dome. Punched micas are used in the assembly in order to locate the various parts and connecting wires; barium getters are also included. The separate parts were all subjected to high vacuum heat treatment before assembly. The 'diode' is then sealed on to a high vacuum pump system and evacuated to a pressure of approximately 10^{-6} mm Hg, and the glass envelope of the tube outgassed by baking at a temperature of 400°C for one hour. After allowing the tube to cool, the cathode temperature is raised to about 1300°K in order to decompose the barium–strontium carbonate to barium–strontium oxide. When the pressure has fallen again to

10^{-6} mm Hg the anode is outgassed by raising it to red heat, the getters are fired and the tube sealed off from the pump. By using the construction and processes outlined above it is possible to produce the barium–strontium oxide by a method comparable with that used in normal valve production and at the same time make thermojunction probe connections to points within the cathode coating.

§3. Theory

If it is assumed that the barium–strontium oxide coating may be considered as a simple ionic crystal semiconductor with n_σ impurity centres per cm^3 of energy level σ, Fermi level ζ and lower limit of the conduction band χ, it may be shown that the following approximate relationship holds:

$$\chi-\zeta=\tfrac{1}{2}(\chi-\sigma)+\frac{\boldsymbol{k}T}{2\boldsymbol{e}}\log\frac{(CT)^{3/2}}{n_\sigma}$$

where $C=2\pi\boldsymbol{mk}/\boldsymbol{h}^2$, $\chi-\zeta$ is the internal work function ϕ and equals $\frac{1}{2}(\chi-\sigma)$ at 0°K. From this it follows that

$$\kappa=\kappa_0\exp\{-\tfrac{1}{2}\boldsymbol{e}(\chi-\sigma)/\boldsymbol{k}T\},\qquad \kappa_0\propto l\sqrt{n_\sigma}\quad\text{(conductivity)}\quad\ldots\ldots(1)$$

$$\frac{dV}{dT}\simeq\tfrac{1}{2}(\chi-\sigma)\frac{1}{T}+\frac{\boldsymbol{k}}{2\boldsymbol{e}}\log\frac{(CT)^{3/2}}{n_\sigma}-\frac{2\boldsymbol{k}}{\boldsymbol{e}}\quad\text{(thermoelectric power)}\quad\ldots\ldots(2)$$

$$I_s=A_0T^{5/4}\exp\left[-\frac{\boldsymbol{e}}{\boldsymbol{k}T}\{\psi+\tfrac{1}{2}(\chi-\sigma)\}\right],\qquad A_0\propto\sqrt{n_\sigma}$$

$$\text{(total emission current)}\quad\ldots\ldots(3)$$

where ψ is the external work function.

The tube described in the previous section gives facilities for making the following observations: (i) the variation with temperature of conductance $\bar{\kappa}$ between any pair of probes, (ii) the variation of thermoelectric power dV/dT with temperature, (iii) the variation of total emission current $i_s(F=0)$ with temperature, (iv) the variation of anode–cathode conductance $\bar{\kappa}'$ with temperature.

It is seen therefore that determination of the internal work function ϕ, the total work function $\psi+\phi$, the number of impurity centres n_σ and comparison between the internal conductance and anode–cathode conductance is possible.

§4. Methods of Measurement

Since a temperature gradient must exist in the cathode when the heater temperature is raised there will be a potential difference between the tape probes, even when no current is being drawn through the cathode, as a consequence of the very large thermoelectric effects inherent in semiconductors. The conductance was therefore found from the slope of the (V, I) characteristics where V is the potential difference between any pair of probes and I the anode current. These characteristics were found to be straight lines as long as the condition of the cathode did not change during the measurements at any particular temperature. This was checked by frequent re-measurement of the potential difference corresponding to zero current. Furthermore, the conductance is dependent on the state of activation of the oxide, which must be approximately constant over the series of measurements at different temperatures. Measurements of the emission current made at pre-fixed temperature and field conditions were used to estimate the state of activation of the cathode. This condition was checked

after each series of measurements at a particular temperature. The potential difference between the probes was measured by means of a potentiometer, the platinum and 13% rhodium–platinum tapes being short circuited when used as voltage probes. The temperature of the section of the oxide under consideration was determined by means of the platinum and 13% rhodium–platinum thermojunctions defining the section. It will be noticed that the tapes B and C give a true voltage probe measurement (four-terminal) of the conductance, whereas junction A, although not used as a connection in the anode–cathode circuit, includes the platinum-to-oxide contact through which current is flowing.

The thermoelectric power was measured between the probes A and B, the tape C being used as a secondary heater. In this way a graph could be drawn of V_{AB}, the potential difference between the tapes A and B and $T_A - T_B$, the temperature difference between them; the current through the cathode was zero and the average temperature $\frac{1}{2}(T_A + T_B)$ was kept constant. The slope of this graph gave the thermoelectric power dV/dT. The value of dV/dT was determined for a series of values of $\frac{1}{2}(T_A + T_B)$, the state of activation of the cathode being noted as for conductance measurements.

The variation with temperature of the total emission current corresponding to zero field was determined by a method developed by M. F. Holmes (private communication). This is based on direct current measurements at low cathode temperatures and does not employ any sort of pulse technique. The point of divergence from the method described by Metson *et al.* (1952) is that an experimental correction is applied to ensure that the emission measured is not subjected to any accelerating field. The underlying principle depends on the fact that zero field conditions obtain immediately prior to the development of a space-charge potential minimum and that this development can be determined experimentally.

The anode–cathode conductance was determined from the displacement of the 'retarding field' characteristic of log anode current against V_a from its theoretical value. In order to avoid the poisoning of the cathode observed by Metson, the anode voltages used were less than 6 volts, with the result that comparatively low cathode temperature (of the order 800°K and less) was necessary for the measurements.

As in the case of conductance and thermoelectric power measurements, the condition of activation of the cathode was noted after each series of observations at a particular temperature.

Four experimental diodes D, E, F and G were used for these measurements. In two others, J and K, only observations of the total emission current, total conductance (A–C) and anode–cathode conductance were possible owing to a defective tape connection.

§5. Experimental Results

Figure 2 shows the variation of log conductance between the probes B and C as a function of reciprocal absolute temperature. During these measurements the state of activation of the oxide was estimated after each temperature as indicated above. A progressive reduction in this value was observed as the temperature was increased from the points Y to the points Z, after which, with further increase in temperature, the value remained sensibly constant. The value of the conductance between the points Y and Z was then corrected for the

difference in activation from the constant value, the points between X and Z resulting. Before correction these points suggest that the curves consist of two branches of different slopes and hence different work functions. After correction, however, the points lie on a single straight line. This upholds the supposition that only measurements made on the oxide at an approximately constant level of activation could be included in a series of observations. This condition has been adhered to in all the results shown.

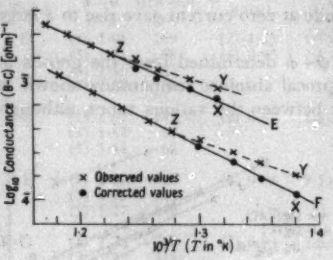


Figure 2.

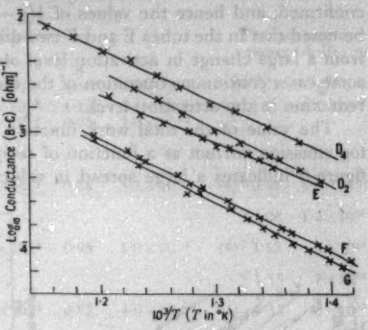


Figure 3.

Figure 3 shows the variation of log conductance (B–C) as a function of the reciprocal of absolute temperature for the four tubes D, E, F and G, there being a change in the general state of activation between the readings D_2 and D_4. The range of temperature over which these measurements were possible was restricted by the conditions of the level of activation imposed above. It will be seen that these graphs are all straight lines, confirming the relationship $\bar{\kappa}=\bar{\kappa}_0 \exp(-e\phi/kT)$ or, since the dimensions of the cathode coating may be taken as constant, $\kappa=\kappa_0 \exp(-e\phi/kT)$ where κ_0 and ϕ are constants defined above.

Figure 4 shows the variation of log conductance (A–B) with the reciprocal absolute temperature. Some of these curves show a 'knee' at low temperatures.

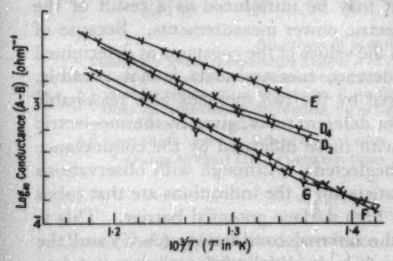


Figure 4.

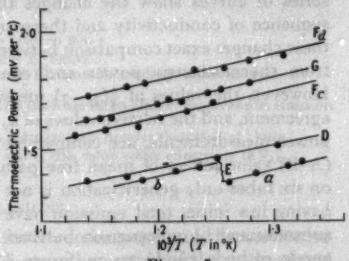


Figure 5.

Since the slope of these lines for temperatures greater than that corresponding to the 'knee' is the same as for the conductance (B–C) lines it must be concluded that these 'knees' are not characteristic of the oxide matrix, but are a result of

the inclusion of the oxide–metal contact in the section of the oxide on which measurements are being made.

It was found that the direction of the thermoelectric e.m.f. was such as to make the metal positive with respect to the oxide at the hot junction, giving confirmation that barium–strontium oxide is an excess semiconductor. Figure 5 shows the variation of thermoelectric power as a function of reciprocal absolute temperature. It is seen from these curves that the relationship $dV/dT \propto 1/T + \text{constant}$ is confirmed, and hence the values of $\frac{1}{2}(\chi - \sigma)$ and n_σ may be calculated. It will be noted that in the tubes E and F two distinct lines may be drawn; this results from a large change in activation level observed during the measurements. In some cases continuous operation of the diode at zero current gave rise to a large reduction in the activation level.

The value of the total work function $\psi + \phi$ determined from the graphs of log emission current as a function of reciprocal absolute temperature shown in figure 6 indicates a large spread in value between the various tubes, although

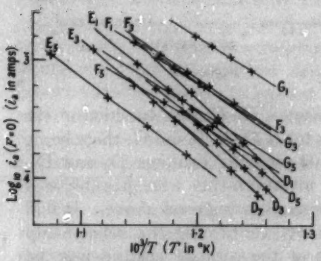


Figure 6.

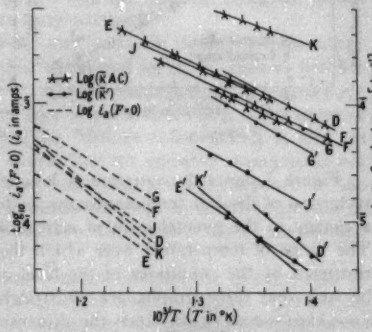


Figure 7.

the internal work function was found to be of the same order. Further, these series of curves show the changes that may be introduced as a result of the sequence of conductivity and thermoelectric power measurements. Because of these changes exact comparison between the values of the constants as determined from thermoelectric power and conductance measurements is not possible. However, the values of $\frac{1}{2}(\chi - \sigma)$ measured by the two methods give reasonable agreement, and the relative values of n_σ for different tubes, given by thermoelectric power measurements, are comparable with those indicated by the conductance (A–B) value (effect of mean free path neglected). Although with observations on six tubes only generalization is not satisfactory, the indications are that tubes having low initial total emission have a high surface potential barrier. This is substantiated by comparison between the internal conductance (A–C) and the anode–cathode conductance (figure 7). A high initial total emission and low total work function is associated with comparable values of the two conductances of comparable work function, whereas a low total emission and high total work function is associated with widely differing values of the two conductances (that of the anode–cathode conductance of the order of ten times conductance A–C)

of different work functions, that of the anode–cathode conductance being comparable with the total work function.

§6. Summary of Experimental Results

The following is a summary of the experimental results; the sequence of measurements is shown by the numbers in brackets.

	Total emission current		Conductance AB		Conductance BC		Thermoelectric power (A–B)	
Tube	$\psi+\phi$ (ev)	$i_a(F=0)$ (μA)	ϕ (ev)	$\bar{\kappa}_{AB}$ (Ω^{-1})	ϕ (ev)	$\bar{\kappa}_{BC}$ (Ω^{-1})	ϕ (ev)	n_σ
D	(1) 1·81	169	(2) 1·13	$1\cdot5\times10^{-3}$	1·13	$2\cdot4\times10^{-3}$	(6) 1·1	$8\cdot5\times10^{16}$
	(3) 1·81	98	(4) 0·95	$2\cdot1\times10^{-3}$	0·95	$3\cdot0\times10^{-3}$		
	(5) 1·55	127						
	(7) 1·67	80						
E	(1) 1·61	141	(2) 0·88	$2\cdot8\times10^{-3}$	0·88	$1\cdot3\times10^{-3}$	(4) [a] 0·99	$6\cdot0\times10^{16}$
	(3) 1·58	87						
	(5) 1·44	62					[b] 1·08	$1\cdot2\times10^{16}$
F	(1) 1·21	325	(2) 0·98	$9\cdot5\times10^{-4}$	0·98	$5\cdot0\times10^{-4}$	(4) [c] 1·17	5×10^{15}
	(3) 1·33	322						
	(5) 1·27	143					[d] 1·15	7×10^{15}
G	(1) 1·29	760	(2) 0·94	$7\cdot6\times10^{-4}$	0·93	$4\cdot0\times10^{-4}$	(4) 1·15	8×10^{14}
	(3) 1·29	316						
	(5) 1·33							

	Total emission		Conductance AC		Anode–cathode conductance	
Tube	$\psi+\phi$ (ev)	$i_a(F=0)$ (μA)	ϕ (ev)	$\bar{\kappa}_{AC}(\Omega^{-1})$	ϕ (ev)	$\bar{\kappa}'$ (Ω^{-1})
D	(7) 1·67	80	(9) 1·03	$4\cdot6\times10^{-4}$	(8) 1·65	$9\cdot2\times10^{-5}$
E	(5) 1·44	62	(7) 0·85	$3\cdot7\times10^{-4}$	(6) 1·32	$3\cdot6\times10^{-5}$
F	(5) 1·27	141	(7) 0·90	$2\cdot8\times10^{-4}$	(6) 0·90	$2\cdot8\times10^{-4}$
G	(5) 1·33	175	(7) 0·99	$2\cdot85\times10^{-4}$	(6) 1·09	$2\cdot6\times10^{-4}$
J	(1) 1·37	160	(4) 0·87	$3\cdot5\times10^{-4}$	(2) 1·05	$8\cdot4\times10^{-5}$
	(3) 1·40	98				
K	(1) 1·67	82	(2) 0·80	*$1\cdot2\times10^{-3}$	(3) 1·62	$5\cdot1\times10^{-5}$
	(4) 1·65	89				

All values of i_a and $\bar{\kappa}$ refer to 800°K.

[a] Values for lower temperature region *a* of curve E, figure 5.
[b] Values for higher temperature region *b* of curve E, figure 5, readings taken after *a*.
[c] Values for temperature increasing, F_c in figure 5.
[d] Values for temperature decreasing, F_d in figure 5 readings taken after *c*; large drop in emission noted between *c* and *d*.
AB lines show ' knee '; values quoted refer to temperatures greater than that of ' knee '.
* Owing to short circuit between tapes B and C this value refers effectively to $\bar{\kappa}_{AB}$.

§7. Conclusions

From the temperature dependence of both conductance and thermoelectric power it may be concluded that the barium–strontium oxide of the oxide-coated cathode within the temperature range 700°K to 900°K behaves as an impurity semiconductor having a single value of the work function.

From the very large difference observed between the internal conductance of the oxide and the total anode–cathode conductance for tubes of low emission,

that is, high total work function, it appears that the absolute level of emission is controlled by the state of the surface layer of the cathode, the internal conductance and internal work function being of the same order for all tubes.

Acknowledgments

The author wishes to express her appreciation to the Engineer-in-Chief of the General Post Office for permission to publish the information contained in this paper, and to Mr. F. T. Turner for his skilled technical assistance.

References

Hannay, H. B., McNair, D., and White, A. H., 1949, *J. Appl. Phys.*, **20**, 669.
Loojes, R., and Vink, H. J., 1949, *Philips Res. Rep.*, **4**, 449.
Metson, G. H., Wagener, S., Holmes, M. F., and Child, M. R., 1952, *Proc. Instn Elect. Engrs*, Pt. III, **99**, 69.
Metson, G. H., *et al.*, 1949, *Nature, Lond.*, **164**, 589.
Young, J. R., 1952, *J. Appl. Phys.*, **23**, 1129.

Viscometry—The Meniscus Resistance in Capillary Flow of Liquids

By G. F. N. CALDERWOOD and E. W. J. MARDLES
Royal Aircraft Establishment, Farnborough, Hants

MS. received 18*th December* 1953

Abstract. Experiments first described by Guy Barr in a paper to the Society on surface tension effects in viscometry, relating to the influence of a wetting agent on flow times in standard viscometers have been continued. A disturbing surface tension effect has been traced to meniscus resistance, particularly of the recessional meniscus, the resistance being decreased by wetting agents and increased with increase of contact angle. With a variety of oils of different chemical composition and of different surface charactisterics the reproducibility of viscosity results obtained with standard viscometers and by the procedure laid down in the latest I.P.-A.S.T.M. method varied up to 0·5%, a coefficient of error greater than that usually obtained with mineral oils alone.

§ 1. Introduction

In view of the adoption since 1st July 1953 of freshly distilled water at 20·00°c with viscosity of 1·0038 centistokes as sole calibrating liquid, surface tension effects on time of flow in the standard viscometers become important when a high degree of precision is required.

To minimize the error arising from the high surface tension of water compared with that of oils a master viscometer is specified. The master viscometer has a capillary length of 40 cm compared with 10 or 12 cm of the routine type and is employed for determining accurately the viscosity of secondary calibrating liquids. The routine viscometers for testing oils are calibrated with the secondary calibrating liquids, usually hydrocarbons in the form of white spirits or oils of relatively low surface tension, and not with water, otherwise an error due to surface tension arises.

In a paper to the Physical Society (1946) on errors in viscometry due to surface tension, Guy Barr found that the calculated corrections for two liquids, with surface tensions differing rather more than those of water and benzene, diverge by 0·5% when a mean head of 10 cm is assumed and a simple experiment demonstrated that this estimate is probably low owing to effects of drainage on the shape of the meniscus. The simple experiment referred to was the effect on the flow-time of water in a B.S. No. 0 viscometer of adding 0·02 ml. of a solution of a wetting agent to the upper level, raised to 1 cm above the upper timing mark. The time of flow was reduced from 1348 to 1334 sec, approximately 1% reduction. The experiment of touching the upper meniscus with a rod dipped in a solution of a wetting agent has been made by the authors using standard viscometers and the original result of Barr has been confirmed and found also with water–glycerol mixtures. Additional experiments have been made in which an index of liquid has been allowed to move down a tilted tube under a known constant force in order to elucidate the phenomenon and it has been concluded from the results of the latter set of experiments that the transitory film of liquid

clinging to the walls in the wake of the moving liquid column forms part of the meniscus acting as a variable and elastic resistance. The advancing meniscus appears also to possess friction, but to a smaller extent than that of the receding meniscus.

§2. Experiments with Standard Viscometers and I.P. Master Viscometers (both U-tube and suspended level)

Using the viscometer in the standard way the effect of touching the top meniscus with a rod dipped in a wetting agent was to reduce the flow time in seconds for water in a B.S. No. 1 viscometer by an average amount of 0·7%.

With the master viscometers an average reduction of 0·15% was observed. The solution of wetting agent was dyed so that any diffusion into the capillary could be observed. A little diffusion occurred in all cases.

Subsequent tests showed that the amount of dyed wetting agent was sufficient to increase the flow time up to 0·2% when allowed to disperse throughout the whole volume of water present. Several varieties of wetting agents were used and they all gave similar effects. The effect of touching the meniscus in the lower bulb of the U-tube viscometer with wetting agent was usually to increase the time of flow slightly but by an amount within 0·2% reproducibility. The effect of touching the upper mensicus with a wetting agent in the case of a glycerol–water mixture of viscosity 100 cs in a B.S. No. 3 viscometer was to decrease the time of flow from 313·6 sec to 311·2 sec, a reduction of 0·75%. Some tests were made to ascertain whether the flow times of an oil were likely to be affected by coating the upper bulb with a thin film of an oleophobic substance, such as sodium or lithium stearate, when an increased contact angle would be caused. It was found that an increased time of flow of the order 0·4% was usually obtained. When the lower bulb was coated, any effect observed was considerably smaller and when both bulbs were coated the differences found were not increased beyond 0·4%.

When the upper bulb of the master viscometer was coated with a hydrophobic film of 1 volume of oleic acid, 1 volume of mineral oil by rinsing with a 20% solution in chloroform, draining for two hours and drying while the viscometer was inverted to avoid entry into the capillary, the change in flow time for water at 20°c (average of three runs) was from 496·8 sec to 498 sec, a percentage increase of 0·24. When both bulbs were coated the flow time was 498·2 seconds.

§3. The Resistance caused by Menisci

Although it is generally assumed that the Jamin effect, namely the capacity of a chain of droplets in a tube to sustain a pressure, is non-existent on clean surfaces or with organic liquids (Smith and Crane 1930, Jamin 1860), the presence of a measurable meniscus resistance found by the authors with all liquids examined in scrupulously cleaned tubes is of some importance to the general theory and practice of viscometry. Water has a relatively high meniscus resistance and may reach several dynes per cm of meniscus periphery in clean glass tubes and this value increases with time of drainage or by contamination of the surface by impurities from the atmosphere. With organic liquids such as *n*-hexane, chloroform or benzene, a meniscus resistance of the order up to a dyn cm^{-1} of meniscus periphery was found (Calderwood *et al.* 1953).

This meniscus resistance contributing to higher viscosity values is quite apart from the effect of reversal of flow at the meniscus which also contributes to higher viscosity values (Calderwood *et al.* 1953). Barr ascribed the anomalous surface tension effects to the influence of drainage on the meniscus shape. Our experiments indicate that the resistance of the upper receding meniscus does vary with degree of drainage and cleanliness of the surface. The meniscus resistance of oils appears to be much smaller than for water but the resistance is measurable and is greater the greater the contact angle, as indicated by results obtained with coating the walls of a tube by an oleophobic film of soap, in the same way that the meniscus resistance of water is increased on hydrophobic surfaces.

The transitory wet film above the receding meniscus apparently forms part of the meniscus and its resistance is both elastic and rhythmically variable depending presumably on the length, thickness and stability etc. of the film.

The menisci were watched during slow motion under a microscope, a little uranine dye having been added to the water. They were oblate-spheroidal in shape and apparently with zero or very small contact angles. Their peripheries were irregular and from the advancing meniscus movement began at tiny points where a tongue of liquid moved forward, especially if the tube was tapped. The tongue had a contact angle or ridge around it and appeared to emerge either between the periphery and the glass or through a weak point in the periphery of the meniscus which behaved as if attached to an adsorbed film ahead. The meniscus in recession was attached to the film of liquid left behind on the walls not only to an immediately adjacent film showing interference colours but also to a thicker zone further behind ending with a fringe of filiform tongues of liquid. The movement of the meniscus and the fringe was jerky and synchronous.

Meniscus resistance and explanations of the versatile behaviour of liquid indices are to be sought in the properties of the thin films adjacent to and attached to menisci.

§4. Rate of Movement of Liquid Indices down Inclined Tubes under Constant Force

These experiments were made to ascertain the actual resistance in dyn cm^{-1} of meniscus periphery. With freshly distilled water in a tube of 0·18 cm diameter, cleaned with chromic acid mixture, rinsed and dried after about one hour's standing, the force in dynes required to begin perceptible motion, observed with a microscope in the absence of appreciable vibration, was of the order 10 dynes. The initial movement was jerky and under the microscope it could be seen that the periphery was irregular. Taking into account the irregular periphery the Jamin resistance is approximately 5 dyn cm^{-1}.

By increasing the angle of tilt of the tube the force is increased in proportion to the sine of the angle and a new rate of movement can be observed. It was found convenient for smooth adjustment of the angle to use a disc tilted at any convenient angle. The tube was fixed diametrically and horizontally on the tilted disc which could be smoothly rotated to allow the tube to be inclined from the horizontal at any small angle required.

Some results are shown in figure 1 which show the effect on rate of movement of touching the rear meniscus of an index of water with a wetting agent. Although touching the advancing meniscus resulted in only a slightly lower resistance, a

strong effect was observed when the rear meniscus was treated with a wetting agent, the index being immediately released and the graph relating dynes with rate of movement being shifted nearly to the origin. On the other hand, by treating the walls of the tube with a hydrophobic film of oil or wax the graph is shifted considerably in the other direction. The same phenomenon was observed with a double index of water with silicone fluid or an oil, the only effective change being when the rear meniscus of the water was touched with wetting agent.

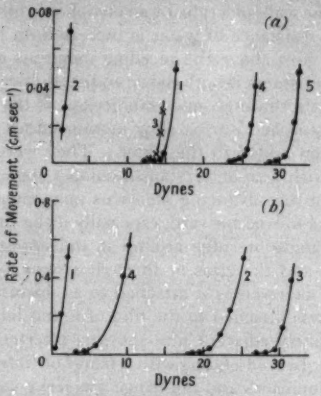


Figure 1. The effect on rate of movement of touching the rear meniscus of an index of water in a tube with a wetting agent ; internal diameter of tube 0·18 cm, length of index 5 cm.

(*a*) 1. Freshly distilled water in clean dry tube.
2. Freshly distilled water in clean dry tube rear meniscus touched with wetting agent.
3. Freshly distilled water in clean dry tube front meniscus touched with wetting agent.
4. Freshly distilled water in tube contaminated with trace of mineral oil.
5. Freshly distilled water in tube coated with thin film of paraffin wax.

(*b*) 1. Silicone fluid (20 cs viscosity) in clean glass tube.
2. Index of 2·5 cm silicone fluid and 2·5 cm water (water in front) in clean glass tube.
3. Index of 2·5 cm silicone fluid and 2·5 cm water (water in rear).
4. Index of 2·5 cm silicone fluid and 2·5 cm water (water in rear) rear meniscus touched with wetting agent.

In table 1 are given results obtained with a mineral oil using the same method. Additional results showing the effect of a standard anti-spreading composition are given in figure 2.

Table 1. The rate of Movement of an Index of 10 cm Length of a Mineral Oil of Viscosity 169 cs in Tube of 0·18 cm Diameter

Force (dynes)		0·4	0·82	1·07	1·63	2·00	2·37	2·72	3·05
Rate of movement (10^{-3} cm sec^{-1})	Clean	2·5	3·3	4·4	7·6	12·9	210·5	400·0	617
	Coated*	0·03	0·33	0·83	1·28	—	2·8	5·2	12·2

* With oleophobic film of lithium stearate.

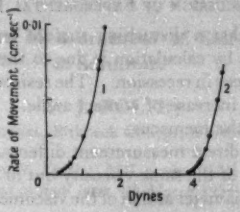


Figure 2. The rate of movement of a 10 cm length index of mineral oil (viscosity 17·8 cs) in a tube, internal diameter 0·0768 cm before and after coating with a film of ferric and cobalt stearates (antispreading composition DTD. 644).

1. In clean tube.
2. In tube coated with metal stearates.

§5. Elastic Recoil of an Index after Movement

It was observed that if the tube was placed carefully in a horizontal position after movement, the index returned more or less completely to its original position, depending on the extent and rate of movement. The results given in table 2 were obtained by allowing the index of liquid to flow down the tube for a few centimetres and by reversing the tilt until the index was restored to its original position. Then, when the tube was placed carefully in a horizontal position, the liquid began to move back along the wetted portion until it reached the dry walls when movement ceased.

Table 2. The Elastic Recoil of a Liquid Index 5 cm in length

l=return movement in cm, v=rate of recession, 10^{-2} cm sec^{-1}

(*a*) *Formamide*							
l	0·2	0·4	0·6	0·9	1·5	2·5	3
v	7	4	3	2	1·2	0·8	0·63
lv	1·4	1·6	1·8	1·8	1·8	2·0	1·9
(*b*) *Formamide* 1 *vol.*, *water* 1 *vol.*							
l	0·1	0·2	0·3	0·4	0·5	0·6	0·7
v	18	12	8·2	4·5	2·6	1·5	1·2
lv	1·8	2·4	2·5	1·8	1·3	0·9	0·8
(*c*) *Water base hydraulic fluid*							
l	0·25	0·5	1	1·5	2		
v	3·3	2·2	1·3	1·05	0·8		
lv	0·83	1·1	1·3	1·6	1·6		

The rates of movement in table 2 were obtained from the slope of tangents to the graph relating distance of recession and time taken. The elastic nature of the recoil is indicated by the product lv which usually varies but little over the length of recoil.

Further evidence on the elastic nature of the film of liquid in the wake of the liquid index was afforded during slow movement by the rhythmic character of the flow of the index. The graphs relating distance moved and time, and of the time taken to move 1 mm at different points along the tube were wavy throughout.

§6. Discussion of Experimental Results

The results indicate that a disturbing surface tension effect in viscometry greater than that obtained by calculation is due to the resistance of the menisci, especially that of the meniscus in recession. The resistance is decreased by wetting agents and increased with increase of contact angle. It is also variable with the degree of drainage above the meniscus.

Attempts to detect by direct measurement differences in the rate of drainage of various kinds of liquids of the same viscosity but of different chemical composition in tubes of the same diameter as that of the viscometer bulb were inconclusive. By using standard B.S. viscometers carefully calibrated as laid down in the standard method and used in the standard I.P.–A.S.T.M. way with a variety of liquids including mineral oils, oils of sebacate ester base, silicone fluids, water base hydraulic fluids, etc., it was found that the reproducibility of results varied up to 0·5% of the mean value, a value greater than that required by the standard method for oils alone, indicating some disturbing influence, presumably due to surface tension and drainage effects.

Acknowledgment

Acknowledgment is made to the Chief Scientist, Ministry of Supply, for permission to publish.

References

Barr, G., 1946, *Proc. Phys. Soc.*, **58**, 575.

Calderwood, G. F. N., Douglas, H. W., and Mardles, E. W. J., 1953, *Analyt. Chem.*, **25**, 835.

Jamin, J., 1860, *C. R. Acad. Sci., Paris*, **50**, 172.

Smith, W. O., and Crane, M. D., 1930, *J. Amer. Chem. Soc.*, **52**, 1345.

A Micro-Method for the Measurement of Self-Diffusion Coefficients of Solutions*

By J. ROLFE†
Department of Physics, Birkbeck College, University of London

Communicated by R. Fürth ; MS. received 15th October 1953
and in final form 9th March 1954

Abstract. A modification of the micro-diffusion method of Fürth and Zuber for measuring the diffusion of a β-active electrolytic solution into a non-radioactive but otherwise identical solution is described. The apparatus can be used to determine self-diffusion coefficients of ions and their dependence on concentration. Results on sodium iodide, using iodine 131, and orthophosphoric acid, using phosphorus 32, are given and discussed on the basis of existing theories. Further applications of the method are suggested.

§1. Introduction : Survey of Previous Work

The micro-methods of Fürth (1925) and Zuber (1929) for the measurement of diffusion in liquids have been extensively used (see, for example, Fürth 1945) for the determination of the dependence of the diffusion coefficients of electrolytic solutions on concentration, and the results were found to be in general agreement with the theories of Onsager and Fuoss (1932) and Sitte (1932). Since then the theory has been further developed (Onsager 1945, Harned 1947), and recent exact experiments by Harned and collaborators (1945, 1947), using another method, have given agreement with theory to within 1%.

The theory was also worked out for the self-diffusion of ions in solutions, and in view of the importance of testing the theory it was suggested to the author by Dr. R. Fürth that a modification of the above-mentioned micro-method should be worked out so that diffusion of radioactive isotope tracers in non-radioactive solutions of the same electrolyte could be measured. Practically this is a self-diffusion process.

Since, in a self-diffusion experiment, the total concentration throughout the diffusion system remains constant, the self-diffusion coefficient obtained is a 'differential' diffusion coefficient, as opposed to the 'integral' diffusion coefficient often quoted in measurements of the ordinary diffusion coefficient, where measurements are taken over an appreciable range of concentration. (It is, of course, the 'differential' diffusion coefficient which is required to find the variation of diffusion coefficient with concentration.) Methods such as that of the porous diaphragm diffusion cell, and the method of chemical analysis of layers of a diffusion column, which for ordinary diffusion give 'integral' coefficients, can therefore be used to give 'differential' self-diffusion coefficients. A general discussion of 'differential' and 'integral' diffusion coefficients is given in Jost (1952).

Some earlier work on the self-diffusion of monovalent ions using radioactive tracers will be briefly mentioned here. Adamson (1947) and Adamson *et al.*

* Substance of the author's Ph.D. thesis, University of London, 1953.

† Now at Research Laboratory, Associated Electrical Industries, Ltd., Aldermaston, Berkshire.

(1949) used the porous diaphragm method to measure self-diffusion of Na^+ ions in NaI solutions ; in the second paper the disadvantages of this method are well illustrated. The results, though reproducible to 1%, are 7% higher than the theoretical values, and it was found that 'surface diffusion' of adsorbed ions was taking place in the pores of the membrane. Whiteway *et al.* (1950), using a porous diaphragm to investigate the self-diffusion of Ag^+ ions in $AgNO_3$, obtained agreement with the theoretical value at zero concentration, but disagreement with theory at higher concentrations. Wang and Kennedy (1950) used the layer analysis method to measure the self-diffusion of Na^+ and I^- in NaI solutions, and obtained results with an error of 1%. While the values they obtained for I^- agreed with theory, values for Na^+ were 7% higher, as in Adamson's experiments. Wang and Kennedy's experiments seem the most reliable to date, since calibration of the apparatus with solutions of known diffusion constants is not necessary, as in porous diaphragm methods, and no surface adsorption of ions can take place.

§2. The Fürth Micro-Method

An outline only of the method is given here. Further details and references will be found in the review article by Fürth (1945) and in the original papers by Fürth (1925), Ullman (1927), Zuber (1932), and Fürth and Zuber (1934). Diffusion coefficients in liquids being of the order of 10^{-5} cm^2 sec^{-1}, the time needed for a diffusion experiment in a cell of dimensions of about one centimetre is about 10^5 sec, or 27 hours, and since most diffusion cells are larger than this (Longsworth 1945), and the time of experiment is proportional to the square of the length of the diffusion cell, most diffusion experiments take times which are of the order of days, during which time the apparatus must be kept at a very constant temperature to avoid the disturbing effects of convection currents. In the 'microdiffusiometer' diffusion is observed over lengths of about 2 mm, corresponding to a time of experiment of about an hour. In such a short time no special precautions need be taken to keep the temperature constant.

The lower half of a microdiffusion cell is filled with a solution of concentration c_0, and the upper half of the cell with pure solvent. At time $t=0$ diffusion is started by pulling out the slider separating the two solutions, and the upward progress of a concentration c_0/n ($n>2$) is followed using, in the case of a coloured solution, a slit in a microscope objective. If a series of observations is taken of x, the distance above the slider where a concentration c_0/n occurs, and the time t when x is observed, then a graph of x^2 against t will be a straight line, irrespective of whether the diffusion coefficient D is constant or a function of the concentration c. When D is independent of c it can be calculated from the slope of the straight line. With the given initial conditions, assuming an infinite diffusion cell, the solution of Fick's second law is

$$x^2/t = D[2\,\mathrm{erf}^{-1}(1-2/n)]^2 \qquad \ldots\ldots(1)$$

where

$$\mathrm{erf}\,x = (2/\sqrt{\pi})\int_0^x \exp(-\xi^2)\,d\xi \qquad \ldots\ldots(2)$$

and erf^{-1} is the inverse error function. In a single experiment a family of x^2/t lines can be obtained by following the progress of several concentrations, and an average value of D obtained. Even if D depends on concentration, a suitable analysis of a family of x^2/t curves will give the function $D(c)$. Details of calculations necessary are given by Zuber (1932). For self-diffusion, of course, the simple solution (1) is valid.

§ 3. Apparatus

The apparatus consists of a movable diffusion cell, with one face transparent to β-radiation, and a fixed collimating slit with Geiger counter and counting-rate meter measuring the radioactivity, and hence the concentration, of any horizontal section of the liquid in the diffusion cell.

Figure 1 shows the diffusion cell used. The back plate supporting the cell is a standard 3 in. × 1 in. microscope slide, and the sides of the cell are formed from glass of the same thickness, cemented to the back of the cell with 'Araldite'. The front face of the cell is formed by a one inch square of mica, 0·05 mm thick, attached to the rest of the cell with shellac. Two sliders are used, the lower to separate the two solutions before diffusion and the upper, kept closed during experiments, to confine any changes in concentration caused by evaporation to the region above the top slider, where there will be no effect on the diffusion. These sliders are made of spring steel, 0·005 in. thick and 1 in. long, ground to fit exactly between the front and back walls of the cell. The lower slider, when closed, fits into a very small slot on the opposite side of the cell, to give it greater rigidity when the upper half of the cell is being filled.

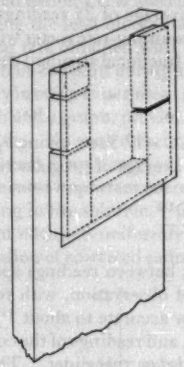


Figure 1. Diffusion cell. For clarity the sliders are omitted.

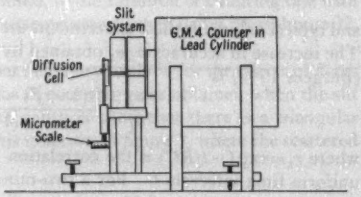


Figure 2. The scanning apparatus.

The diffusion cell is mounted in a scanning apparatus (figure 2). The lower part of the back of the diffusion cell is held between vertical guides, and its vertical movement controlled by a micrometer adjusting screw, so that the height of the cell can be adjusted with an error of 0·01mm. The slit system, the walls of which have to be thick enough to absorb any γ-rays which may be given off, is formed by four lead blocks separated by spacers, the slit nearer the cell acting as the defining slit with a smaller slit width than the second slit, which was added later for increased γ-ray absorption. By varying the number of spacers the defining slit-width can be adjusted in 0·005 in. steps.

Radiation through the slits is measured with a G.E.C. type G.M.4 Geiger counter, mounted in a 2 in. thick lead cylinder, and a Panax model 40A counting-rate meter. It was found experimentally, over the concentration ranges used in

these experiments, that the counting rate is directly proportional to the concentration of radioactive tracer.

§4. Experimental Technique

By means of small pipettes the lower half of the cell is filled with a radioactive solution of a known concentration, the lower slider closed, and excess solution removed from the upper half with pipettes and strips of filter paper. The upper half of the cell is filled with non-radioactive solution of the same concentration, and the cell placed in the scanning apparatus. Since it is found that radioactive solutions of very high specific activity (20 mc/ml.) are needed, very small quantities (0·15 ml.) of the required solutions are made up by weighing solutions delivered from a micro-pipette into a weighing bottle. Exactly similar non-radioactive solutions are made up using volumes 100 times greater, which are measured with ordinary pipettes.

The initial counting rate from the undiffused solution in the lower half of the cell is determined for distances of 1·0 to 4·0 mm below the slider, in 0·5 mm steps. Variations of the counting rate shown by these readings, apart from that due to β-ray scattering (see below), will detect disturbances due to leakage of the slider, etc. Counting rates are determined by taking the average of 25 readings at the rate of one every five seconds. A single reading of a counting-rate meter with time constant RC, for a counting rate of n per second, has a fractional standard deviation (Schiff and Evans 1936)

$$\sigma_1 = (2nRC)^{-1/2} \qquad \ldots\ldots(3)$$

and typical values for these experiments are $n = 100$, $RC = 12{\cdot}5$ sec, giving $\sigma_1 = 2\%$. The increase in accuracy σ_s/σ_1 obtained by taking an average of s regular readings (Burgess 1949) is

$$\frac{\sigma_s}{\sigma_1} = \frac{[s(1-r_1^2) - 2r_1(1-r_1^s)]^{1/2}}{s(1-r_1)} \qquad \ldots\ldots(4)$$

where $r_1 = \exp(-t/RC)$ is the correlation coefficient between readings spaced at uniform time intervals t. For a two-minute time of observation, with readings every 5 seconds, $\sigma_1/\sigma_s = 0{\cdot}44$, so that readings are now accurate to about 1%.

Diffusion is started by pulling out the lower slider, and readings of the counting rate are taken at various times t and distances x' *below* the slider. The self-diffusion coefficient D is then calculated from the formula obtained from (1). This formula is valid now because c, the total concentration, is constant :

$$D = \frac{x'^2}{4t}[\mathrm{erf}^{-1}(2r-1)]^{-2} \qquad \ldots\ldots(5)$$

where r is the ratio of the counting rate obtained at (x', t) to the initial counting rate, allowing for corrections (see below). Because of the sluggish response, and the fluctuations of the counting-rate meter, the method of following a particular concentration cannot be used, so that x' and t are fixed for the experiments by calculating a time-table of x' and t corresponding to constant ratios of x'^2/t and finding the concentration, and hence D, for each reading. From 13 to 18 readings are taken in each experiment, t varying from 9 to 132 minutes and x' from 0·75 to 4·15 mm. An average value of D is found for each experiment.

The only precautions necessary to avoid temperature gradients during the experiment are that the front face of the cell is not touched with the fingers, and

that the apparatus is surrounded by a tin box to avoid draughts on the cell. The progress of diffusion in the cell seems to be quite insensitive to mechanical vibration, and no special precautions are taken to avoid it.

§ 5. Corrections

All results for D are reduced to a standard temperature of 18°c with a multiplying factor proportional to η/T, where T is the absolute temperature and η the viscosity of water at temperature T. This is a correct procedure for very low concentration, where the self-diffusion coefficient is directly proportional to the ionic equivalent conductivity, for a range of temperatures of a few degrees near 18°c, and was assumed to hold for higher concentrations.

Corrections for the effect of a finite slit width are made by calculating the actual positions χ (not at x', the centre of the slit) at which the observed concentrations occur, for values of x' and t given by the time-table, and taking average values for D, since the correction depends only slightly on D. This correction decreases with increasing t, and is only appreciable for the first six readings when the largest slit width is used.

When iodine 131 is used in experiments, a correction is necessary for the γ-rays traversing the lead blocks forming the slit walls and affecting the counter. The counting rate obtained with the slit closed is subtracted from all readings. It is found to be constant during an experiment. To avoid direct determination of this quantity in every experiment it is calculated from the counting rate of the undiffused solution and a graph, previously determined, of the variation of counting rate with slit width. No such correction is of course necessary when using phosphorus 32, which emits β-rays only.

The most important correction is that necessary to allow for the effects of β-ray scattering in the solution. Observations of counting rates obtained when the slit scans an enclosed uniformly radioactive solution show that there is a triangular distribution of scattered radiation. This is shown in figure 3, where the scattered

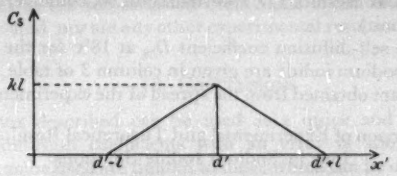


Figure 3. Assumed scattering distribution.

radiation C_s is plotted against distance x' along the cell, when the slit is at a position $x'=d'$. The slope of the sides of the triangle is k, and l is the maximum scattering length. For phosphorus 32, $l=3{\cdot}25$ mm, $k=2{\cdot}50$, and for iodine 131, $l=2{\cdot}5$ mm, $k=5{\cdot}965$. Thus β-scattering in phosphorus 32 solutions takes place over a greater distance than in iodine 131 solutions, but the proportion of scattered radiation in iodine 131 solutions is higher than in phosphorus 32 solutions. This is to be expected, since the maximum energies of β-rays from iodine 131 and phosphorus 32 are 0·6 Mev and 1·7 Mev respectively, and the higher the energy of a β-particle the greater its penetrating power and the smaller the angular deflection on scattering.

The scattering correction for diffusion experiments can be evaluated now by multiplying this scattering function described above by the diffusion function, describing the change of concentration by diffusion after given time, and the integral of the result will give the correction to be applied to each reading. The correction is formulated in terms of integrals of the complementary error functions (Hartree 1935), the general form of the correction being

$$s(d,a)=\frac{ka^2}{2}\left[\text{ii erfc}\frac{d-l}{a}+\text{ii erfc}\frac{d+l}{a}-2\text{ii erfc}\frac{d}{a}\right] \quad \ldots\ldots(6)$$

where $a=2(Dt)^{1/2}$ and k and l have the meanings given above.

The scattering correction is applied by subtracting the product of s and the initial counting rate (before the start of the diffusion) from the readings obtained at distance d and time t.

The application of the correction becomes essential when the slit width is small compared with the maximum scattering length l. This was indeed the case in the experiments with phosphorus 32 solutions, but in the experiments with iodine 131 the slit width was large compared with l and it was not necessary to apply the correction.

§ 6. Experimental Results

The small solid angle subtended by the slit means that solutions of high specific activity (approximately 20 mc/ml.) have to be used. This requirement limits the range of isotopes which can be used to those which can be obtained carrier-free. Furthermore, the γ-radiation of the isotope should have an energy less than 1 MeV to avoid the necessity of a very long collimating slit. At the time the only carrier-free isotopes obtainable from A.E.R.E., Harwell, with a suitable β- and γ-radiation were iodine 131 and phosphorus 32. To avoid chemical manipulations they were used in the chemical form supplied, i.e. as sodium iodide and orthophosphoric acid solutions respectively. (A larger range of high-activity isotopes are now available from the Radiochemical Centre, Amersham. Pure β-emitters with a long half-life, such as caesium 137 and strontium 90, could very easily be used in similar investigations.)

Results for the self-diffusion coefficient D_{18} at 18°C for the iodide ion in five concentrations of sodium iodide are given in column 3 of table 1. The standard deviations quoted are obtained from the spread of the experimental results.

Table 1. Comparison of Experimental and Theoretical Results of D_{18} for Iodide Ions in Sodium Iodide Solutions

c (formula wts. per litre)	$D_{18}\times10^5$ (cm^2 sec^{-1})		
	Calculated	Rolfe	Wang and Kennedy
0	1·70	1·73 ± 0·025	1·69 ± 0·02
0·03	1·62	1·58 ± 0·08	1·57 ± 0·03
0·058	1·59	1·67 ± 0·03	1·56 ± 0·02
0·175	—	1·71 ± 0·06	1·53 ± 0·02
0·200	—	1·70 ± 0·06	1·53 ± 0·02

The result for infinite dilution ($c=0$) is obtained from experiments where carrier-free isotope solution is diffused into distilled water, and is the mean of five experiments. The other results are obtained from single experiments. (Each experiment consists of about 13 determinations of D.) The error in c is estimated to be 1%.

Values obtained for the self-diffusion coefficient of anions in orthophosphoric acid are given below in table 2. In this case it is impossible to be more precise as to the character of the diffusing ion, since the ions PO_4^{3-}, HPO_4^{2-} and $H_2PO_4^-$ are present in different proportions, varying with the pH of the solution.

Table 2. Self-Diffusion Coefficients of Anions in Orthophosphoric Acid Solutions

c (formula wts. per litre)	$D_{18} \times 10^5$ ($cm^2\ sec^{-1}$)
0	$0{\cdot}60 \pm 0{\cdot}01$
0·029	$0{\cdot}63 \pm 0{\cdot}03$
0·080	$0{\cdot}58 \pm 0{\cdot}03$
0·477	$0{\cdot}55 \pm 0{\cdot}02$
0·737	$0{\cdot}52 \pm 0{\cdot}02$

§7. Discussion

A limiting equation for the self-diffusion coefficient of iodide ions in dilute sodium iodide solutions has been calculated by Gosting and Harned (1951) from the Onsager–Fuoss theory for the concentration range $\sqrt{c}=0$ to $\sqrt{c}=0{\cdot}3$, namely

$$D_{18} \times 10^5 = 1{\cdot}70 - 0{\cdot}438\sqrt{c}\ cm^2\ sec^{-1}. \qquad \ldots\ldots (7)$$

Values of D calculated from (7), in the indicated range of validity, are shown in column 2 of table 1, and the experimental results of Wang and Kennedy (1950), corrected to 18°c, in column 4.

It can be seen that the experimental results obtained by both methods are in general agreement with theory, though the accuracy of the results is not high enough for any further conclusions to be drawn.

No data exist to find the equivalent conductivity of the anions in orthophosphoric acid, since no transport numbers are available. Therefore no theoretical values of the self-diffusion coefficients of anions in orthophosphoric acid can be calculated, nor are any other experimental results available.

§8. Conclusions

The apparatus described can be used as a quick and reliable method of measuring self-diffusion coefficients with an error of about 2%, though the isotopes which can be used are limited to those which can be obtained carrier-free, because of the low sensitivity of the apparatus. For the method to be useful for other isotopes the sensitivity would have to be increased about 1000-fold, and it is difficult to see how this could be carried out except by increasing the size of the apparatus, in which case all the advantages of a micro-method are lost. However, the accuracy could probably be increased to 1% by formulating a slightly more refined scattering correction, taking account of the effect of finite slit width.

Using carrier-free isotopes, further experiments could be performed using this apparatus on self-diffusion in molecular solutions, where the radioactive atom is made part of a larger molecule, and on 'tracer diffusion' (Wang 1952), where a non-common radioactive ion present in small concentration diffuses in a solution of otherwise constant concentration.

Acknowledgments

It is a pleasure to thank Dr. R. Fürth for his constant help and advice throughout the investigation. Thanks are also due to Professor J. D. Bernal for permission to work in the Laboratories at Birkbeck College, to the Department of Scientific and Industrial Research for a maintenance grant, and to the Central Research Fund of the University of London for a grant for the apparatus and radioactive isotopes.

References

Adamson, A. W., 1947, *J. Chem. Phys.*, **15**, 762.
Adamson, A. W., Cobble, J. W., and Nielson, J. M., 1949, *J. Chem. Phys.*, **17**, 740.
Burgess, R. E., 1949, *Rev. Sci. Instrum.*, **20**, 964.
Fürth, R., 1925, *Phys. Z.*, **26**, 719; 1945, *J. Sci. Instrum.*, **22**, 61.
Fürth, R., and Zuber, R., 1934, *Z. Phys.*, **91**, 617.
Gosting, L. J., and Harned, H. S., 1951, *J. Amer. Chem. Soc.*, **73**, 159.
Harned, H. S., 1947, *Chem. Rev.*, **40**, 461.
Harned, H. S., and French, D. M., 1945, *Ann. N.Y. Acad. Sci.*, **46**, 267.
Harned, H. S., and Nuttall, R. L., 1947, *J. Amer. Chem. Soc.*, **69**, 736.
Hartree, D. R., 1935, *Mem. Manch. Lit. Phil. Soc.*, **80**, 85.
Jost, W., 1952, *Diffusion in Solids, Liquids and Gases* (New York: Academic Press, Inc.).
Longsworth, L. G., 1945, *Ann. N.Y. Acad. Sci.*, **46**, 211.
Onsager, L., 1945, *Ann. N.Y. Acad. Sci.*, **46**, 241.
Onsager, L., and Fuoss, R. M., 1932, *J. Phys. Chem.*, **36**, 2689.
Schiff, L. I., and Evans, R. D., 1936, *Rev. Sci. Instrum.*, **7**, 456.
Sitte, K., 1932, *Z. Phys.*, **79**, 320.
Ullman, E., 1927, *Z. Phys.*, **41**, 301.
Wang, J. H., 1952, *J. Amer. Chem. Soc.*, **74**, 1182.
Wang, J. H., and Kennedy, J. W., 1950, *J. Amer. Chem. Soc.*, **72**, 2080.
Whiteway, S. G., MacLennan, D. F., and Coffin, C. C., 1950, *J. Chem. Phys.*, **18**, 229.
Zuber, R., 1929, *Phys. Z.*, **30**, 882; 1932, *Z. Phys.*, **79**, 291.

Measurement of the Surface Properties of Germanium

By J. B. GUNN*
Elliott Brothers (London) Limited, Research Laboratories, Borehamwood, Herts.

Communicated by A. E. De Barr ; MS. received 10th December 1953, and in amended form 17th February 1954

Abstract. New techniques are described for the measurement of the surface properties of n-type germanium. These make use of the properties of a new type of rectifying contact. Some preliminary results are given, which are interpreted as demonstrating the existence of a p-type surface layer separated from the interior of the specimen by a rectifying barrier. The observed properties of the surface layer do not support existing theories of rectification at a point contact, and are found to depend sensitively on surface condition.

Two new types of transistor are described whose mechanism depends on the presence of the surface layer.

§1. Introduction

The theoretical implications of the existence of additional energy levels at the surface of a semiconductor are now well known (Bardeen 1947). The presence of acceptor levels on an n-type specimen may give rise to a p-type layer separated from the interior by a p–n junction, which will be a good rectifier provided holes and electrons cannot recombine too freely either near the surface or in the interior of the specimen. Qualitative considerations similar to those discussed by Shockley (1949) in connection with the 'abrupt transition' p–n junction are expected to apply to this case. Under conditions of small inverse bias such a surface structure should behave as a conducting 'skin' surrounding the specimen and separated from it by the nearly insulating junction, across which only the constant saturation current flows. With increasing bias the thickness and transverse conductance of the skin should decrease, and at large biases some sort of Zener breakdown current is to be expected like that observed in solid junctions.

It is proposed to introduce the terms *external contact* to describe a relatively low-resistance electrical contact made to the surface layer without disturbing its properties, and *internal contact* to describe one which, by destroying the rectifying junction, enables connection to be made to the interior of the semiconductor.

Attempts have been made to apply the theories of external contacts to the properties of germanium point contact diodes and transistors. In these it has generally been assumed that the metal whisker is a true external contact, although its properties are considered to be modified by self-heating, radial symmetry of the current flow, etc. This approach has been successful when applied to the problem of carrier injection (Bardeen and Pfann 1950), but attempts to obtain quantitative agreement with observed current–voltage characteristics have failed, the predicted inverse current being much too small and its activation energy too large (Bennett and Hunter 1951). Aigrain (1952) has sought to improve

* Now at Radar Research Establishment, St. Andrew's Road, Great Malvern, Worcs.

matters by taking account of the lateral conductivity of the surface layer, but is forced to assume conductivities so high that they can be accounted for only by assuming that acceptor impurities actually enter into the germanium lattice by diffusion, which seems unlikely in view of the mild surface treatments which may be employed.

Various methods for making large area contacts have been described, which apparently lead to external contacts in varying degrees of approximation. Bray (1950) has described preliminary experiments on several types of rectifying contact. These included rhodium plating, colloidal graphite, and fired-on silver paste, which for various reasons would now be expected to lead to rather poor external contacts.

The properties of evaporated gold contacts have been described by Pantchechnikoff (1952) and Granville and Henisch (1952). The results of the former suggest that true external contacts have been made, but, since they describe a turnover voltage, the contacts made by the latter workers would seem to be of a different type from those to be described here.

In this paper a number of new types of external contact are described, and the results of some preliminary measurements of the surface properties of germanium are given.

§2. Experimental Technique

Material

Mono-crystalline n-type germanium was used in which the hole lifetime was greater than 100 microseconds.

Internal Contacts

Non-rectifying connections were usually made by soldering, but where this was not possible the same effect was obtained by using separate current and voltage leads pressed into contact with the specimen. It was verified that effects due to floating potentials and hole injection by the first electrode were negligible.

Surface Treatment

As a last step before the application of the external contact, the specimen was etched, rinsed in distilled water, blotted with clean filter-paper, and allowed to dry. It was hoped in this way to obtain a clean and reproducible surface, but it was found, towards the end of the work described, that even the careful rinsing technique employed did not eliminate the effect of impurities in the etch. Consistent results were obtained by etching either in 6% hydrogen peroxide (Analar) at 60°c or in CP4 prepared from Analar materials (Heidenreich 1952). In both cases the purity of the etch is believed to have been sufficient for the surfaces produced to be characteristic of clean germanium, contaminated no doubt by water and air. Such surfaces will be referred to as of 'type 0'.

The use of a commercial grade of hydrogen peroxide, containing as little as 0·1% of impurities, was found to produce significantly different surfaces which will be described as 'type 1'. It was with these that most of the data were obtained, and the differences from the type 0 surfaces appear to result from the presence of traces of sodium carbonate. The major impurity was sodium phosphate, which is apparently inactive.

Further treatment of the surface after etching was carried out in only a few ('type 2') cases (§3).

External Contacts

Rectification was first observed at electroplated copper contacts, to which connection could be made either mechanically or by soldering at 80°c. Normal soldering completely destroyed the rectification. Similar contacts of indium have recently been described by Bradley (1953).

The most versatile type of contact has been found to be a conducting paint prepared from silver dust with the addition of a small quantity of an apparently inert polystyrene binder dissolved in xylene or benzene. It was found possible to apply quite complicated contact systems with a fine brush, and contact was made to these with a fine wire carrying a blob of the wet paint. Alternatively, where heavy currents were involved, a thin brass or copper ribbon was painted on one side and then pressed down on to the germanium.

When it was desired to remove and replace the contact many times, to allow gases to diffuse to and from the germanium surface, or to prepare contacts with a standard area, the silver dust was used alone. For this purpose the dust was placed in a hole in a block of Perspex and pressed against a germanium disc, which closed one end of the hole, by means of a steel plunger inserted from the other end. Small holes drilled through the Perspex put the external atmosphere in communication with the dust, to which electrical connection was made by means of the plunger. The diameter of the hole, and hence of the contact area, was 0·125 inch. Such 'dry' contacts were also prepared using a nickel dust and a silver dust composed of nearly spherical particles, but no significant difference from the normal type was observed. This was also the case when the actual external contact was formed by a layer of colloidal graphite applied to the germanium before the silver dust was pressed into contact.

A third type of external contact has been prepared using an aqueous electrolyte as the contacting material. This was either absorbed in a disc of filter-paper, which was used to replace the silver dust in the 'dry' holder, or else the germanium disc was sealed with W-wax to the open end of a glass tube, into which the liquid was introduced through a side arm. Connection to the liquid was then made with a small platinum disc. The properties were again generally similar to those of the normal 'dry' contacts.

Measurement of Electrical Properties

In some cases the current–voltage characteristics of the specimens were measured statically. In other cases the characteristics were traced out at 50 c/s on a cathode-ray tube, and measurements made from photographs of the trace. The circuits supplying current to the specimen were arranged so that forward and inverse currents were independently variable, and so that there was a short period during which neither circuit supplied current. The spot consequently remained stationary at the true origin during this time and a bright spot appeared on the record. Small zero errors in the d.c. amplifiers feeding the cathode-ray tube could thus be allowed for.

§ 3. Rectification Effects

Rectification phenomena have been observed in every external contact made to the surface of n-type specimens, but only low-resistance, non-rectifying characteristics have been observed on p-type material.

A characteristic typical of those obtained with n-type material is shown in figure 1. The sections OA, OB, and BD will be termed the *forward current*, *saturation current* and *breakdown current* regions respectively.

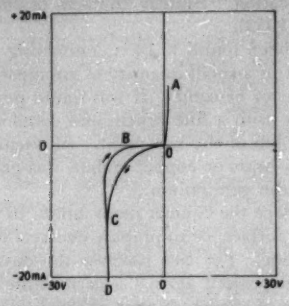


Figure 1. Typical oscilloscope trace of the current voltage characteristic of an external contact (' dry ' contact, type 1 surface, 6 Ω cm material).

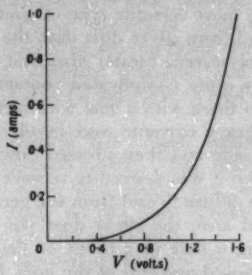


Figure 2. Static forward characteristic of a low resistance painted contact.

Forward Current

In the forward current region the resistance was always found to be low, and often determined mainly by the bulk resistance of the germanium. A specimen was constructed from 6 ohm cm material, using a painted contact to a type 1 surface, in which this effect was minimized by the geometry. The contact had an area of about 0·06 cm², and had the forward characteristic shown in figure 2. Between 0·1 and 1·0 A this curve has a value of $d(\ln I)/dV$ of about $3{\cdot}4\,v^{-1}$; somewhat higher values are obtained at lower current densities.

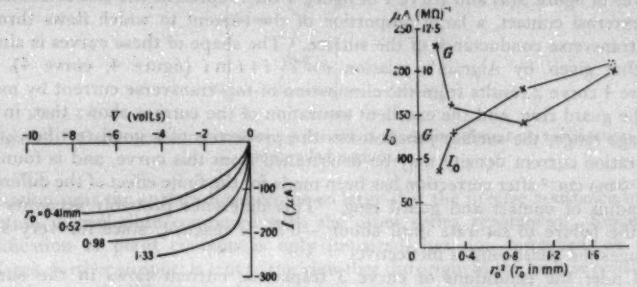
Several transistor-like specimens were constructed, consisting of a thin germanium disc with painted external contacts on either side, and measurements were made of the current gain α when these were used as emitter and collector. A lower limit for the injection ratio γ of the emitter could be deduced from this, since γ must be greater than α for a structure such as this with a collector multiplication factor of unity. Values of α from 0·5 to 0·85 were found for plates about 0·05 cm thick.

There is some evidence to suggest that α, and hence presumably γ, depended on surface condition, but in general the forward characteristic was found to be fairly reproducible and much less dependent on surface preparation than was the inverse characteristic.

Saturation Current

Although the hysteresis loop in the inverse part of figure 1 appears to extend into the saturation current region, it is in fact apparently associated with changes in the voltage at which the breakdown current begins to flow. Since this voltage increases with time during the application of inverse voltage, the static method usually permits measurements to be made over a wider range than with the oscilloscope, without interference from the breakdown current.

Figure 3 shows some characteristics obtained from a silver paint contact on 6 ohm cm material (type 1 surface). After each curve had been taken, the mean radius r_0 of the approximately circular contact was measured optically, and the area was then increased by adding more paint. These curves may be analysed in terms of a small conductance G in parallel with an ideal rectifier which does not, however, reach its saturation current I_0 until much higher voltages than those predicted by any existing theories of rectification. This is essentially what is done in most studies of point contacts, and, if contacts of the present type were supposed to consist merely of many independent points, electrically in parallel on a uniform surface without transverse conductance, one would expect to obtain straight lines passing through the origin on plotting I_0 and G against r_0^2. This is by no means the case, as is shown in figure 3 (*b*).



(*a*) (*b*)

Figure 3. (*a*) Static inverse characteristics of painted contacts of varying radii r_0.
(*b*) Variation of I_0 and G with contact area.

Figure 3 can be explained by assuming (*a*) that the surface was inhomogeneous, and the initial contact happened to be made on a part having a higher-than-average saturation current density, or (*b*) that the *total* available inverse current was limited and consisted largely of holes which were generated mainly at one point and were able to diffuse over distances comparable with the contact dimensions, or (*c*) that the contact was truly external and that a surface conductance contributed to the 'saturation current', as discussed by Aigrain. In order to resolve these possibilities the specimen was cleaned and re-etched, and a painted contact was applied of mean radius 0·60 mm. This had the characteristic shown as curve 1 of figure 4. An annular contact, of inner radius 0·93 mm, was then applied as nearly concentrically as possible, and further measurements were

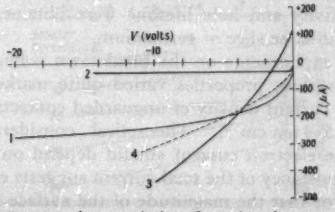


Figure 4. Static inverse characteristics of a painted contact with guard ring.

made of the current I_1 to the contact as a function of the potentials V_1 and V_2 of the contact and annulus respectively.

When V_1 equalled V_2 (i.e. annulus used as guard ring) curve 2 was obtained. The difference between this and curve 1 shows that (*a*) is incorrect. The form of curve 3, which was taken with $V_2 = -1{\cdot}5$ v, eliminates (*b*) also, since on this hypothesis curve 3 should lie between curve 1 and the V_1 axis. Further measurement showed that, under these conditions, the changes in I_1 were accompanied by equal and opposite changes in the current to the annulus.

When the specimen was examined again, a day later, it was found that the influence of the guard ring had greatly decreased, and currents only slightly larger than those of curve 2 were obtained in all cases.

All the foregoing results are qualitatively explicable on assumption (*c*). The curves of figure 3 (*a*) and curve 1 of figure 4 then represent the characteristics of an external contact, a large proportion of the current to which flows through the transverse conductance of the surface. The shape of these curves is similar to that given by Aigrain's relation $v = 1 - i + i \ln i$ (figure 4, curve 4). In figure 4 curve 2 results from the elimination of this transverse current by means of the guard ring, and the excellent saturation of the current shows that, in this voltage range, the surface junction has the properties of a good rectifier. The saturation current density may be determined from this curve, and is found to be $1{\cdot}6$ mA cm^{-2} after correction has been made for the finite effect of the difference in radius of contact and guard ring. This difference probably accounts also for the failure to saturate until about $-0{\cdot}3$ v is reached, since for very small voltages the guard ring is ineffective.

Under the conditions of curve 3 transverse current flows in the surface layer, but only in the space between the annulus and the contact. This has the effect of placing a conductance between the two, which accounts for the extension of this curve beyond the region bounded by the voltage axis and curve 1, and also for the changes in annulus current. This conductance is evidently non-linear, however.

The change in properties is taken to indicate that the surface conductance decreases with the passage of time, and this interpretation is subtantiated by later observations (§ 4).

It is difficult to give meaningful figures for the magnitude of the saturation current, since it was markedly dependent on both temperature and illumination. The effect of the latter was eliminated by making all measurements in the dark, but the temperature was, in this work, known only to be near that of the room. The value of $1{\cdot}6$ mA cm^{-2} obtained for the specimen of figure 4 may be compared with the value of $0{\cdot}5$ mA cm^{-2} calculated on Shockley's theory with the assumption that electron–hole pairs are excited only in the body of the material. Values for the resistivity and hole lifetime were obtained from a filamentary specimen cut from the same slice of germanium.

In the course of experiments on the breakdown voltage of dry contacts, in which some of the surface properties varied quite markedly, it was observed that the 'saturation' current density of unguarded contacts departed only rarely from the range $2{\cdot}5$–4 mA cm^{-2}. Theoretical considerations (Gunn 1954) suggest that only the electron current should depend on the condition of the surface, so that the constancy of the total current suggests either that the electron current is negligible or that the magnitude of the surface potential rise remains constant.

Abrasion with fine emery paper was found to render the surface completely non-rectifying, and subsequent etching reduced the inverse current continuously to its normal value, as shown in figure 5. Together with the fact that such a surface has a very high recombination velocity, this result suggests that the rate of

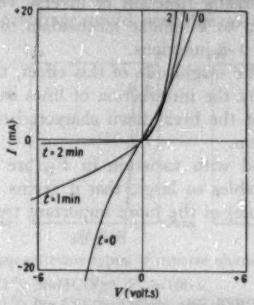


Figure 5. Oscilloscope traces showing the effect of etching time t on rectification at an abraded surface.

excitation near the surface has become so large that the inverse resistance of the junction is negligible compared with the other series resistances. Since the rectification of point contacts is only impaired, but not destroyed, on such surfaces, further support is lent to the view that the origin of the inverse resistance is fundamentally different in the two cases.

In the case of liquid contacts most electrolytes were found to have no effect on the inverse current, but aqueous solutions of salts of zinc and mercury had the effect of increasing the saturation current by an amount, sometimes quite large, which depended on the average forward and inverse currents. Similar effects were observed when distilled water was used as the electrolyte but diffusion restricted by the use of filter-paper or of a very closely spaced ($\sim 10^{-4}$ cm) platinum electrode.

Very large changes in the saturation current were, of course, caused by the injection of holes from an emitter contact. Currents of 100 mA have been obtained in this way from a dry contact, with good saturation and a sharp 'knee'. Painted contacts generally became very noisy at quite low current densities. The collector characteristics of a typical low-power transistor are shown in figure 6.

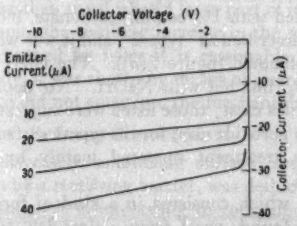


Figure 6. Static collector characteristics of a coaxial transistor using external contacts.

Breakdown Current

The difference between the inverse current at any point on a characteristic and the extrapolated value of the saturation current (which may include a surface component) will be termed the breakdown current. It is, beyond a certain voltage, a very rapidly rising function of inverse voltage (figure 1), and has tentatively been ascribed to a similar mechanism to that responsible for the 'Zener' current in solid p–n junctions.

In order to specify the magnitude of this effect, the breakdown voltage V_B was arbitrarily defined by the intersection of lines extrapolated back from the relatively straight part of the breakdown characteristic, and forward from the 'saturation current' region.

The effects associated with variation in V_B are so complicated, and the interaction between variables so large, that it seems worth while to give only a brief summary (see table) of the more important trends.

Variable x	$\partial V_B/\partial x$	Remarks
1. Resistivity :		type 0 surface
(*a*) 6–20 ohm cm	~ 0	
(*b*) unknown, but $\ll 6$ ohm cm	$+$	
2. Concentration of Na_2CO_3 :		
(*a*) in etch (H_2O_2)	$+$	0·1% as impurity
(*b*) in rinse	$++$	1% in distilled water
3. Concentration of H_3PO_4 :		
(*a*) in etch (H_2O_2)	~ 0	pH 3
(*b*) in rinse	$--$	pH 3
4. Time after preparation		
(*a*) type 0 surface	~ 0	
(*b*) type 1 surface	$-$	times of a few days involved
(*c*) type 2 surface	$++$	times of a few days involved
5. Flow of inverse current :		
(*a*) hole	~ 0	time dependent :
(*b*) breakdown	$+$	time dependent : ' time constant '=0·1–10 sec
6. Oxidation at up to 600°c in air	$-$	
7. Composition and pH of electrolytic contact	~ 0	
8. Presence of water vapour	enhances 5 (*b*)	
9. Temperature :		
(*a*) type 0 and 1 surfaces	~ 0	
(*b*) type 2 surface	$-$	

Of the effects given, 2 and 5 appear to be the most important. Surfaces which had been treated with 1% sodium carbonate, blotted and allowed to dry without rinsing, are described as 'type 2', and it was with these that the highest values of V_B were obtained (figure 7(*a*)). The effect was obtained also with $NaHCO_3$ and H_2CO_3, but not with NaOH. No doubt other chemical treatments equally efficacious exist; those listed were suggested by the analysis of the commercial hydrogen peroxide used for the type 1 surfaces. There was evidence to suggest that this treatment operated mainly by greatly increasing the magnitude of effect 5.

The latter effect, which consisted in a kind of forming by the passage of inverse current, was found, apart from a time lag, to be entirely reversible.

There seemed always to be a limiting value of V_B which could not be exceeded, whatever the peak current, but it was often possible to observe the rising part of the (V_B–breakdown current) curve (figure 8). This effect serves to explain the existence of regions of apparent negative resistance in the breakdown characteristic as in figure 2. Since the inverse current was supplied from a high impedance source, the voltage could continue to rise while the current

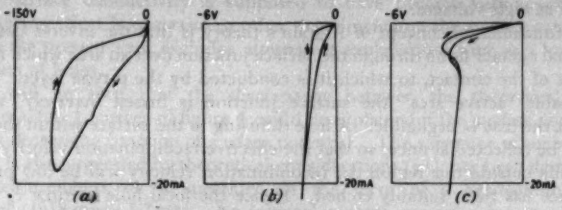


Figure 7. Some extreme types of inverse characteristic.
(*a*) Surface treated with 1% $NaHCO_3$. V_B=100 v.
(*b*) Surface treated with 1% Na_2CO_3 and slight excess of H_3PO_4. V_B=1 v.
(*c*) Type 0 surface showing true negative resistance.

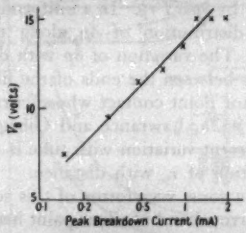


Figure 8. Variation of V_B with peak breakdown current (type 1 surface, 6 Ω cm, stored *in vacuo* for 20 hours).

began to fall. Isolated cases of negative resistance did occur in which the effect could not be invoked as an explanation, but they were rare, and the region of negative resistance was usually small compared with the peak voltage (figure 7 (*c*)).

The distinction between hole and breakdown currents in effect 5 has been made on the basis of experiments in which the saturation current was increased by injecting holes from an emitter while the breakdown characteristic was observed. It was found that the effect of a given peak value of the inverse current in raising V_B was much smaller than usual when a large part of it consisted of holes introduced in this way, despite the resulting increase in the mean value. If the breakdown current is not carried by holes it is reasonable to assume that it consists of electrons. Finally, it was shown by means of guard ring experiments that the surface conductance did not enter into the mechanism of breakdown.

§ 4. Surface Conductivity

The existence of a transversely conducting surface layer, separated from the interior of the specimen by a rectifying barrier, was deduced from the curves of figure 4. In this connection it is interesting to compare curve 1 with that calculated on Aigrain's theory, which should apply in this case. Values of

1·6 mA cm^{-2} for the saturation current density, and $4{\cdot}3\times10^{-6}$ ohm^{-1} (at $V=-1{\cdot}5$ v) for the surface conductance, may be obtained from the contact dimensions and curves 2 and 3. When these values are substituted in the relation $v=1-i+i\ln i$ between reduced voltage v and reduced current i, the broken curve 4 is obtained. Although qualitatively of the right shape and magnitude, this curve predicts a current which is too small at low voltages and too large at high voltages.

The fundamental concept of Aigrain's theory is that the inverse current to an external contact flows through the surface junction over an area which is larger than that of the contact, to which it is conducted by the p-type layer. Within this variable 'active area' the surface junction is biased inversely, whereas outside it the bias is negligible. A hole diffusing to the surface within the active area will be collected at once, so that the effective recombination velocity is very high, while outside this region the recombination velocity will be low provided the surface has been suitably etched. Hence the local hole lifetime τ_p will be greatly reduced by the presence of the active area.

The illumination of germanium by light falling within the photoconductive band results in an increase δp in the hole density, which is limited by an equilibrium between the rate I of quantum absorption and the rate of recombination, and is given by $\delta p=I\tau_p$. In a uniformly illuminated germanium filament the equilibrium distribution of δp along the length will thus be proportional to that of τ_p. The variation of δp with distance can be measured by applying a voltage pulse between the ends of the filament and sweeping the holes along it past a collector point contact whose current is modulated by the local hole density (Gunn 1952a, Lawrance and Gibson 1952). A cathode-ray tube display of collector current variation with time is thus also an approximate representation of the variation of τ_p with distance.

In figure 9(*a*) are shown some waveforms of this sort from a freshly etched filament around which a narrow band of silver paint had been applied at the end

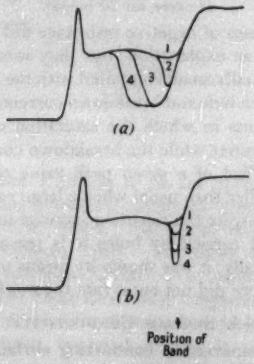


Figure 9. Sketches of typical waveforms showing increase of active area. (*a*) Fresh type 1 surface. Active area increases from 1 (band voltage=0) to 4 (maximum band voltage). (*b*) Same surface after ageing. No change in active area from 1 to 4.

opposite to the collector. For curve 1 no bias was applied to this band, and a nearly uniform τ_p is observed (the initial spike and the final rise are artefacts). For curves 2 to 4 increasing amounts of inverse bias were applied, and the active area can be seen to extend for some distance along the specimen. Figure 9 (*b*) shows the behaviour of the same specimen a few hours later. The active area no longer appeared to increase with increasing bias, a result which can be interpreted if the surface conductivity is supposed to have become very much smaller. This ageing effect is observed in other experiments on the surface conductance, and in the present case excludes alternative explanations such as a longitudinal diffusion of holes.

It will be seen that the discrepancy between the theoretical, 4, and experimental, 1, curves of figure 4 could be explained if the surface conductance were a decreasing function of surface potential. The existence of such an effect, which is also suggested by theoretical considerations (§ 1), has been demonstrated by measurements on a specimen in the form of a germanium filament 0·7 mm × 0·18 mm × 3·2 mm. An internal contact was provided at one end by soldering, and three external contacts in the form of bands about 0·3 mm wide were painted round the filament near the other end (figure 10). The conductance

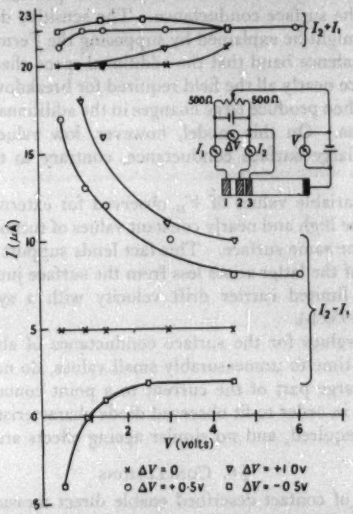


Figure 10. Variation of surface conductance with surface potential.

of the strip of surface (width about 0·29 mm) between bands 1 and 2 was measured using the circuit shown; band 3 was employed as a guard ring to prevent unwanted leakage from band 2 to base. Since the active areas of the bands overlapped considerably, the potential of the surface was taken as the mean of those of the bands. The difference between the value of I_2-I_1 for a given value of ΔV and the value (independent of V) for $\Delta V=0$ was taken as the current flowing in the surface conductance, which is thus seen to decrease with surface potential

as expected. The variation in I_2-I_1 cannot be accounted for by curvature of the individual (I, V) characteristics, since I_2+I_1 always remained nearly constant.

The system described is, in principle, capable of providing power amplification, and is similar to the 'unipolar transistor' of Shockley (1952).

§ 5. Discussion

The effects described are in many cases so complex that theoretical discussion of all the results of this preliminary survey would be pointless. The effects are, however, generally reproducible, and certain tentative conclusions may be drawn.

The magnitude of the saturation current of the surface junction, and its relative independence of surface condition, combine to suggest that it is a hole current arising in the interior of the semiconductor and reaching the surface by diffusion. The value of this current density ($\sim$1 mA cm^{-2}) is significantly different from the density of the 'saturation' current in point contacts ($\sim$1 A cm^{-2}). Since the results given in the table suggest that the breakdown current is of a different kind at the semiconductor surface, it is probably an electron current. In this case it cannot originate from Zener effect in the semiconductor, but must in some way arise outside the surface, perhaps in a film of impurity. Further evidence against the internal Zener mechanism is provided by the correlation between V_B and the surface conductance. The sensitive dependence of V_B on surface condition might be explained by supposing the Fermi level at the surface to lie so near the valence band that the additional space charge due to holes was sufficient to produce nearly all the field required for breakdown; small changes in Fermi level could then produce large changes in the additional field to be supplied by the applied bias. On this model, however, low values of V_B should be associated with a large surface conductance, contrary to the results observed during ageing.

The low and variable values of V_B observed for external contacts stand in sharp contrast to the high and nearly constant values of turnover voltage for point contacts made to the same surface. This fact lends support to the view that the inverse resistance of the latter arises less from the surface junction than from the combination of a limited carrier drift velocity with a system having radial symmetry (Gunn 1952b).

The observed values for the surface conductance of about 5×10^{-6} ohm^{-1}, and its decay with time to unmeasurably small values, do not support Aigrain's hypothesis that a large part of the current to a point contact is carried by the surface layer, since in order to fit observed diode characteristics a value of about 2×10^{-3} ohm^{-1} is required, and no similar ageing effects are found.

§ 6. Conclusions

The new types of contact described enable direct measurement to be made for the first time of a number of properties of the surface of n-type germanium. The preliminary results given do not support the view that the inverse properties of point contacts can be simply related to those of the surface junction even if the surface conductance is taken into account.

Acknowledgments

I am indebted to the Directors of Elliott Brothers (London) Ltd. for permission to publish this paper, and to Mr. H. Wolfson of Standard Telephones and Cables Ltd., Ilminster, for the germanium used.

References

Aigrain, P., 1952, *Ann. Phys., Paris*, **7**, 140.

Bardeen, J., 1947, *Phys. Rev.*, **71**, 717.

Bardeen, J., and Pfann, W. G., 1950, *Phys. Rev.*, **77**, 401.

Bennett, A. I., and Hunter, L. P., 1951, *Phys. Rev.*, **81**, 152.

Bradley, W. E., 1953, *Proc. Inst. Radio Engrs, N.Y.*, **41**, 1702.

Bray, R., 1950, *Purdue University Quarterly Progress Report*, **8**, 27.

Granville, J. W., and Henisch, H. K., 1952, *Proc. Phys. Soc.* B, **65**, 650.

Gunn, J. B., 1952 a, British Patent Application No. 5413/52 ; 1952 b, *Proc. Phys. Soc.* B, **65**, 908 ; 1954, submitted for publication.

Heidenreich, R. D., 1952, U.S. Patent 2,619,414.

Lawrance, R., and Gibson, A. F., 1952, *Proc. Phys. Soc.* B, **65**, 994.

Pantchechnikoff, J. I., 1952, *Rev. Sci. Instrum.*, **23**, 135..

Shockley, W., 1949, *Bell Syst. Tech. J.*, **28**, 435; 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 1365.

Elastic and Piezoelectric Coefficients of Dipotassium Tartrate (DKT)

By R. BECHMANN† and S. AYERS
Post Office Research Station, Dollis Hill, London

MS. received 8th January 1954

Abstract. A new determination has been made of the elastic and piezoelectric coefficients of dipotassium tartrate (DKT) and of their behaviour with temperature. A table is included comparing the properties of some zero-temperature-coefficient cuts of DKT, EDT and quartz.

§1. Introduction

This is one of a series of investigations into the properties of different water-soluble piezoelectric crystals with a view to their use in electric wave filters. The constants have been determined by the method described by Bechmann (1950).

§2. Crystallographic Data

Dipotassium tartrate ($K_2C_4H_4O_6 \cdot \frac{1}{2}H_2O$) commonly known as DKT belongs to the monoclinic sphenoidal class (C_2). Its growth habit is illustrated in figure 1, together with a diagram of the crystallographic and orthogonal axes related to it. The cleavage planes are 100 and 001. The unit cell dimensions taken from

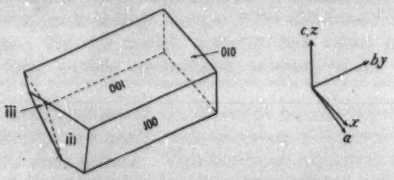


Figure 1. Growth habit of DKT.

Analytical Chemist, 1948, are $a = 20{\cdot}101$ Å, $b = 5{\cdot}049$ Å, $c = 15{\cdot}490$ Å and the angle $\beta = 90° 51'$. (The a and c axes have been interchanged making c the smaller axis in accordance with the text of the I.R.E. Standards, 1949. The choice of axes for DKT in figure 6 of the Standards is not in agreement with the text and corresponds to an older selection.) Mason (1950) kept the a and c axes as originally defined in the *Analytical Chemist* and the relation between his coordinates x', y', z' and the coordinate system x, y, z used here is given by the scheme

	x	y	z	
x'	$-\cos\theta_0$	0	$\sin\theta_0$	
y'	0	-1	0	
z'	$\sin\theta_0$	0	$\cos\theta_0$	.....(1)

where $\theta_0 = 90° 51'$.

§3. Elastic Coefficients

The determination of the elastic coefficients s_{ik} was made in accordance with the methods described by Bechmann (1950, 1952) and the new values at 20°C are given in table 1. Their temperature coefficients Ts_{ik} were measured over

† Now with The Brush Laboratories Co., Cleveland, Ohio.

the range 20°–50°c and are quoted in table 2. Mason's (1946, 1950) values transformed according to the matrix (1) have been included in both tables for comparison. The main coefficients and their temperature coefficients agree substantially but there are large discrepancies for some of the cross coefficients. In calculating the new values, Mason's values for the coefficients of thermal expansion were used and the density was taken as 1·987 g cm^{-3} from the *Analytical Chemist*, 1948.

Table 1. Elastic Coefficients of DKT in 10^{-13} cm^2 dyn^{-1}

s_{ik}	Mason †	New values	s_{ik}	Mason †	New values
s_{11}	38·9	47·5	s_{55}	81·7	102
s_{22}	33·7	35·3	s_{66}	119·2	122·5
s_{33}	22·6	24·0	s_{15}	8·5	−7·5
s_{12}	−10·6	−17·4	s_{25}	−5·4	8·0
s_{13}	−16·4	−8	s_{35}	−6·5	−14·0
s_{23}	−0·7	−6·2	s_{46}	5·5	−6·8
s_{44}	104·3	113·5			

† Mason's values transformed.

Table 2. Temperature Coefficients of the Elastic Coefficients of DKT in 10^{-6}/°c

Ts_{ik}	Mason*	New values	Ts_{ik}	Mason*	New values
Ts_{11}	818	910	Ts_{55}	530	
Ts_{22}	740	775	Ts_{66}	846	895
Ts_{33}	572	535	Ts_{15}	643	−55
Ts_{12}	446	1305	Ts_{25}	−6617	2200
Ts_{13}	1120		Ts_{35}	844	850
Ts_{23}	762	−16	Ts_{46}	−1845	140
Ts_{44}	−62	−135			

* Mason's (1946) values taken from curves and transformed.

§4. Piezoelectric Coefficients

The piezoelectric coefficients were determined by the method described by Bechmann (1950) which has already been used with satisfactory results on EDT and lithium sulphate monohydrate (Bechmann 1950, 1952). However, with DKT although accurate values were derived for most of the coefficients, some discrepancies were found amongst the coefficients which appear in the determination as pairs i.e. $(d_{22}-d_{34})$, $(d_{22}-d_{16})$ and $(d_{16}-d_{34})$. Although measurements in several planes gave consistent values for each individual pair the difference between any two pairs did not agree with the third. Unfortunately it would take too long to explain these difficulties in detail here or to outline all the work done in this connection. Many cross checks have been provided and in each case they only confirm the values for the individual pairs. All the obvious causes of error have been investigated and many possible explanations considered but so far without success. It has therefore been decided to include in the set of piezoelectric coefficients the 'best' values for d_{16} and d_{34} at the date of publication, quoting them with wide confidence limits, in the hope that the discrepancies may be resolved at a later date. The coefficient d_{22} has been determined by a new method involving lengthwise excitation of a bar along the Y axis and since the theory is not yet complete this value cannot be quoted with as high an accuracy as the rest of the coefficients.

The new values are given in table 3 where they are compared with the published values of Mason (1950), Jaffe (1948) and Spitzer (1951). Jaffe used the same coordinate axes as Mason and their values have been transformed using the matrix (1). Spitzer used the same axes as the present work but took Z instead of Y for the symmetry axis; his values have therefore been transformed cyclically to bring them into line with the other sets.

A new set of values for the temperature coefficients of the piezoelectric coefficients in the range 20°–50°C is given in table 4.

Table 3. Piezoelectric Coefficients of DKT in 10^{-8} e.s.u. dyn^{-1}

	Mason †	Jaffe †	Spitzer †	New values
d_{14}	66·3		−62·5	69·0
d_{16}	−28·0		36·4	40 ± 3
d_{21}	10·7	15·0	14·4	16·2
d_{22}	−8·5	−10·2	−14·6	−13·5
d_{23}	1·9	1·2	2·3	2·2
d_{25}	22·2	17·3	18·8	18·9
d_{34}	−7·9		−12·2	−17 ± 3
d_{36}	25·4		−19·7	−24·4
$d_{22}-d_{16}$	19·5		−49·2	−50
$d_{22}-d_{34}$	−0·65		−2·4	0·5
$d_{16}-d_{34}$	20·2		46·8	57·0

† Values transformed.

Table 4. Temperature Coefficients of the Piezoelectric Coefficients of DKT in 10^{-4}/°C

Td_{ik}	New values	Td_{ik}	New values
Td_{14}	−13	Td_{34}	−270 ± 60
Td_{16}	∼−2	Td_{36}	10
Td_{21}	9	$T(d_{22}-d_{16})$	∼0
Td_{22}	4	$T(d_{22}-d_{34})$	−1040
Td_{23}	−59	$T(d_{16}-d_{34})$	−9
Td_{25}	6		

§5. Plates with Zero Temperature Coefficients of Frequency

5.1. *Square Plates vibrating in the Contour–Shear Mode*

Since Ts_{44} has the opposite sign to Ts_{66}, there is a possibility of finding zero temperature coefficients of frequency in plates XYl_{ϕ}, containing the Y axis and rotated about it. The average temperature coefficient of frequency in the range 20°–50°C for the contour–shear mode of these plates is plotted as a function of the angle ϕ in figure 2. From this curve it can be seen that plates $XYl_{20°\,30'}$ and $XYl_{155°}$ have zero temperature coefficients of frequency for this mode. The change of frequency with temperature for these specimens is parabolic with the peak of the parabola at about 30°C. The temperature at the peak of the parabola is a function of the angle of orientation ϕ and the peak can be made to occur at any required temperature by a slight alteration of this angle. The properties of square plates of these orientations have been investigated with a view to their use in filters and are compared in table 5 with similar cuts of EDT and quartz.

$XYl_{20°\,30'}$ is the better of the two DKT cuts and is in many respects superior to the EDT $XYl_{77°}$ cut. DKT, however, has two strong cleavage planes and is more difficult to process than EDT.

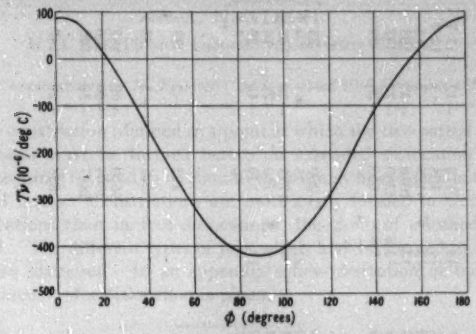


Figure 2. Temperature coefficient of frequency of contour-shear mode of DKT plates XYl_{ϕ}

5.2. *Plates vibrating in the Thickness–Shear Mode*

If there are two angles ϕ at which zero temperature coefficients of frequency occur for the contour–shear mode, there must also be two angles with zero temperature coefficients for the thickness–shear mode. These two angles for DKT are $\phi = 61°$ and $\phi = 115°$. The properties of these cuts are compared in table 5 with those of similar cuts of EDT and quartz. In most respects, the properties of these DKT cuts are less suitable for filters than those of EDT.

5.3. *Rectangular Plates vibrating in the Contour–Extensional Mode*

The properties of DKT rectangular plates, $XYt_{45°}$ and $XYt_{52°\,30'}$, found by Mason (1946) to have zero temperature coefficients of frequency for the contour–extensional mode have been measured and are compared in table 5 (overleaf) with similar cuts of EDT and quartz.

Acknowledgment

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in this paper.

References

Analytical Chemist, 1948, **20**, 491.

Bechmann, R., 1950, *Proc. Phys. Soc.* B, **63**, 577 ; 1952, *Ibid.*, **65**, 368, 375.

Institute of Radio Engineers, New York, 1949, *I.R.E. Standards on ' Piezoelectric Crystals '*, *Proc. Inst. Radio Engrs, N.Y.*, **37**, 1378.

Jaffe, H., 1948, *Phys. Rev.*, **73**, 1467.

Mason, W. P., 1946, *Phys. Rev.*, **70**, 705 ; 1950, *Piezoelectric Crystals and their Application to Ultrasonics* (New York : van Nostrand).

Spitzer, F.,1951, *Arch. Elektr. Übertr.*, **5**, 544.

Table 5. Comparison of the Properties of Zero Temperature Coefficient Cuts of Quartz, EDT and DKT

Material	Orientation	Frequency range (kc/s)	Frequency constant (kc/s mm)	Frequency–Temperature Behaviour: Freq. temp. function	$T\nu^{(1)}$ (10^{-6}/deg)	$T\nu^{(2)}$ (10^{-9}/deg^2)	$\partial T\nu^{(1)}/\partial\phi$ (10^{-6}/deg)	$\partial T_m/\partial\phi$ (°c/deg)	Γ (10^{-6} pF mm^{-1})	TC (10^{-6}/deg)	γ^* (10^{-4})	D (%)
Contour–shear mode of square plates												
Quartz	$ZXl_{-18.5°}$ (CT)	120–500	3080	parab.		−58	4·7	45	118	−12	13·7	96
Quartz	$ZXl_{-38°}$ (DT)	80–300	2073	parab.		−17	−2·2	−74	108	−1	12·4	128
EDT	$XYl_{17°}$	50–150	1150	parab.	†	−730	9·5	−13	125	7	15	133
EDT	$XYl_{77°}$		1160	parab.		−720	−9·3	12	1040	−37	105	135
DKT	$XYl_{22°\ 30'}$		1347	parab.		−780	−6·8	−3·8	4260	−23	342	132
DKT	$XYl_{158°}$		1282	parab.		−540	6·9	5·8	855	−28	89	131
Thickness–shear mode of plates												
Quartz	$ZXl_{-18.5°}$ (AT)	300–10 000	1660·2	cubic		2·5	−4·8		200	2	25	dependent on the thickness–length ratio
Quartz	$ZXl_{-41°}$ (BT)	1000–100 000	2540·5	parab.		−40	2·0	21	75	−8	9·5	
EDT	$XYl_{111°}$	250–1500	923	parab.	†	−720	9·3	4	1795	−30	167	
EDT	$XYl_{164°}$		925	parab.		−500	−9·4	−5	760	−21	76	
DKT	$XYl_{41°}$		1055	parab.		−250	7·0	12	330	−45	29	
DKT	$XYl_{115°}$		1070	parab.			−6·7		40		3·5	
Contour–extensional mode of rectangular plates												
Quartz	$XYt_{18°\ 26'}$ $b/l=0·5$	50–250	2550	linear	−23	—			292	−2	33·5	34
Quartz	$XYt_{-8°}$ $b/l=0·5$		2705	linear	−5	—			312	−5	35·5	33
EDT	YX, $b/l=0·35$	60–150	2040	parab.		−2000			2690	−60	185	12
EDT	YX, $b/l=0·43$		1990	parab.		−2000			2904	−60	200	9
DKT	$XYt_{45°}$ $b/l=0·5$		1723	parab.	†	−860			2847	−30	190	28
DKT	$XYt_{82°\ 30'}$ $b/l=0·5$		1745	parab.		−700			2858	−29	187	28

Notation.

$T\nu^{(1)}$ is first order temperature coefficient of frequency.
$T\nu^{(2)}$ is second order temperature coefficient of frequency.
$\partial T\nu^{(1)}/\partial\phi$ change of $T\nu^{(1)}$ with angle ϕ.
$\partial T_m/\partial\phi$ shift of temperature at peak of parabola with angle ϕ.
$\Gamma = Ct/a$ dynamic capacitance constant.

TC = temperature coefficient of dynamic capacitance.
$\gamma^* = \frac{1}{2}\frac{\nu_p^2 - \nu_s^2}{\nu_s^2} = \frac{1}{2}\frac{C}{C_1}$ = separation between parallel and series resonance.
D = distance to next main mode as percentage of wanted mode.
† as a first approximation zero.

Constrictions and Jogs in Extended Dislocations

By A. N. STROH
H. H. Wills Physical Laboratory, University of Bristol

Communicated by N. F. Mott ; MS. received 10th December 1953

Abstract. A constriction (defined as a point at which the two partial dislocations are recombined) must be formed before an extended dislocation can climb. The energy required is found to be about 6 ev, but may be reduced in the presence of an applied stress. Constrictions are more easily formed in Cottrell–Lomer locked dislocations than in free dislocations; the climb of a locked dislocation is considered. The different types of jogs which may be formed in an extended dislocation are discussed. In an appendix a new derivation of the expression for the line tension of a dislocation is given.

§1. Introduction

The recovery and polygonization of a cold-worked metal is generally considered to be associated with the climb of dislocations out of their glide planes; such climb is possible for perfect dislocations when vacancy diffusion takes place. However, in metals having a face-centred cubic structure, each dislocation is dissociated into two partial dislocations which are separated by a region of stacking fault (Heidenreich and Shockley 1948). The climb of a single partial would result in the formation of an extended surface of lattice imperfection of high energy, so that a necessary preliminary for climb appears to be the recombination of the two partials. If this recombination takes place over a length of only a few atomic spacings, we shall, for brevity, call the resulting form of the dislocation a *constriction*. Constrictions may be produced by thermal fluctuations, possibly aided by an applied stress. The probability of their being so formed will depend on the energy involved, and the calculation of this energy forms the main part of the present paper. We consider both free dislocations and dislocations locked in the manner discussed by Lomer (1951) and Cottrell (1952); we find that constrictions in locked dislocations have the lower energy, but in both cases the energy is too large to permit of their thermal production at room temperatures in the absence of an applied stress. Finally we consider jogs in extended dislocations.

§2. The Energy of a Constriction

We take the undissociated dislocation to lie in the (111) plane and to have Burgers vector $\mathbf{b}=\frac{1}{2}a[1\bar{1}0]$, where a is the atomic spacing; then it will dissociate into two partial dislocations with Burgers vectors $\mathbf{b}_1=\frac{1}{6}a[1\bar{2}1]$ and $\mathbf{b}_2=\frac{1}{6}a[2\bar{1}\bar{1}]$. These two partials are to be re-united at the single point O, so as to form a constriction at this point. We take the origin at the point O and x and y axes such that at a great distance from O the dislocations are parallel to Ox. (No relation is assumed between these axes and the crystallographic axes used to define the Burgers vectors, except that the xy plane is the (111) plane.)

We consider the equilibrium of a short element, of length δs, of the dislocation $\mathbf{b}_1$, at the point P(x, y). The forces on this element, acting in a direction normal to the element, are:

(i) The force due to the dislocation $\mathbf{b}_2$: We assume this to be $A\delta s/2y$, where A is a constant, a form suggested by the expression for the force between two straight dislocations. We will also give A the value appropriate for straight dislocations, namely

$$A = G\mathbf{b}_1\mathbf{b}_2/2\pi = Ga^2/24\pi \qquad \ldots\ldots(1)$$

where G is the rigidity modulus. We have neglected a factor whose value depends on whether the dislocation is of edge, screw or mixed type, and which lies between 1 and $1/(1-\nu)$, where ν is Poisson's ratio.

(ii) Forces due to the interaction of the element with other parts of the dislocation $\mathbf{b}_1$: Mott and Nabarro (1948) have shown that, for a dislocation deviating only slightly from a straight line, these forces may be taken into account by assigning a line tension to the dislocation. We give an alternative treatment of the problem in an appendix, where we show that the restriction that the dislocation should be nearly straight may be replaced by the less stringent condition that the radius of curvature is much greater than r_0, the radius of the core of the dislocation. The line tension of the dislocation is

$$T = \frac{b_1^2 G}{4\pi} \log \frac{R}{r_0} = \frac{a^2 G}{24\pi} \log \frac{R}{r_0} \qquad \ldots\ldots(2)$$

where R is of the order of the length of the dislocation line. The force on the element is then $T(d^2y/ds^2)\delta s$ if the angle between δs and the x axis is not too great.

(iii) Forces due to the stacking fault and to the applied stress: The stacking fault gives rise to a force $F_0\delta s$ tending to reduce the area of the stacking fault, where F_0 is the energy per unit area of the stacking fault, while a constant applied stress σ gives a force $\sigma\mathbf{b}_2\delta s$. Combining these, we have a force $(F_0 + \sigma\mathbf{b}_2)\delta s = F\delta s$, say. If the extended dislocation is in equilibrium, the component of the applied stress parallel to $\mathbf{b}$ must be zero; but a constant stress perpendicular to $\mathbf{b}$ is not of very great interest. A more important case is that in which there is a fluctuating stress parallel to $\mathbf{b}$, with the dislocation at a point where the stress is zero; this is the type of stress met with in the piled-up groups of dislocations considered by Eshelby, Frank and Nabarro (1951). To include such cases we take the force on the element to be $(F + F'y)\delta s$, where F' is proportional to the y derivative of the stress. If the equilibrium of the extended dislocation is stable, F' must be positive.

Equating the sum of the forces obtained in (i), (ii) and (iii) above to zero, we have as the condition of equilibrium of the element

$$\frac{A}{2y} + T\frac{d^2y}{dx^2} - F - F'y = 0. \qquad \ldots\ldots(3)$$

A long way from the constriction $d^2y/dx^2 = 0$ and $y = \frac{1}{2}d$, where d is the width of the extended dislocation. Equation (3) then becomes

$$A/d = F + \tfrac{1}{2}F'd. \qquad \ldots\ldots(4)$$

We introduce now a parameter α such that

$$d = (1-\alpha)A/F. \qquad \ldots\ldots(5)$$

Using equation (4), we may express F' in terms of α:

$$F' = 2\alpha F^2/(1-\alpha)^2 A. \qquad \ldots\ldots(6)$$

As F' varies from 0 to infinity, α varies from 0 to 1; hence we need only consider values of α in this range. Finally we express x and y in terms of dimensionless coordinates u and v:

$$x = \tfrac{1}{2}d(T/A)^{1/2}u, \quad y = \tfrac{1}{2}d(1-v). \qquad \ldots\ldots(7)$$

At a great distance from the constriction $v = 0$.

Equation (3) now reduces to

$$2\frac{d^2v}{du^2} = \frac{1}{1-v} - 1 + \alpha v.$$

This may be integrated at once, giving

$$\left(\frac{dv}{du}\right)^2 = -\log(1-v) - v - \tfrac{1}{2}\alpha v^2. \qquad \ldots\ldots(8)$$

The form of the constriction may be found by integrating equation (8) numerically; the result when $\alpha = 0$ is shown in figure 1. Here we have taken $(\log R/r_0)^{1/2} = 1{\cdot}7$ (see below); a different value would result only in a change of

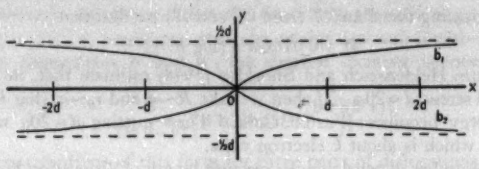


Figure 1.

scale along the x-axis. It is seen from the figure that the inclination of the dislocation lines to the x-axis is small except very near O. The greater part of the energy will be associated with the long lengths of slowly curved dislocations, rather than with the small region of high deformation near O. Hence we regard the approximations made above as justified.

Now we have the following contributions to the energy of the constriction:

(i) the energy of interaction of the two dislocations of amount

$$W_1 = A\int_{-\infty}^{\infty} \log(d/2y)\,dx$$
$$= (AT)^{1/2}d\int_0^1 [-\log(1-v)][-\log(1-v) - v + \tfrac{1}{2}\alpha v^2]^{-1/2}dv$$

on making use of equations (7) and (8);

(ii) the energy due to the line tension

$$W_2 = T\int_{-\infty}^{\infty}\left(\frac{dy}{dx}\right)^2 dx$$
$$= (AT)^{1/2}d\int_0^1 [-\log(1-v) - v + \tfrac{1}{2}\alpha v^2]^{1/2}dv; \qquad \ldots\ldots(9)$$

and (iii) the work, taken with negative sign, done by the applied stress. Since this gives a force per unit length of $F + F'y$, the work done is

$$-W_3 = \int_{-\infty}^{\infty} [F(d-y) + \tfrac{1}{2}F'(d^2 - y^2)]\,dx.$$

Using (5), (6), (7) and (8) we have

$$W_3 = -(AT)^{1/2}d\int_0^1 [v-\tfrac{1}{2}\alpha v^2][-\log(1-v)-v+\tfrac{1}{2}\alpha v^2]^{-1/2}dv.$$

Hence the energy of the constriction is

$$\begin{aligned} W &= W_1 + W_2 + W_3 \\ &= 2(AT)^{1/2}dI(\alpha) \end{aligned} \qquad \ldots\ldots(10)$$

where $$I(\alpha) = \int_0^1 [-\log(1-v)-v+\tfrac{1}{2}\alpha v^2)]^{1/2}dv.$$

Clearly for $0 \leqslant \alpha \leqslant 1$, $I(0) \leqslant I(\alpha) \leqslant I(1)$, and by numerical integration we find $I(0) = 0{\cdot}55$, $I(1) = 0{\cdot}67$, so that $I(\alpha)$ is approximately $0{\cdot}6$ for all α in the range considered. Thus the energy is given by

$$W \simeq 1{\cdot}2(AT)^{1/2}d. \qquad \ldots\ldots(11)$$

We see that W depends on the applied stress only through d, the width of the extended dislocation; thus we should have arrived at the same value for the energy if in equation (3) we had put $F' = 0$ and adjusted F to give the correct value of d.

On substituting for A and T from (1) and (2) we find that

$$W = 0{\cdot}016Ga^2d\,(\log R/r_0)^{1/2}. \qquad \ldots\ldots(12)$$

For aluminium Heidenreich and Shockley (1948) estimate that, in the absence of an applied stress, $d = 20a$. If then we take $R \sim d$ and $r_0 \sim a$, $(\log R/r_0)^{1/2} \simeq 1{\cdot}7$, and the energy becomes $W \simeq 0{\cdot}03Ga^2d$. Then putting $d = 20a$ we find that $W = 0{\cdot}6Ga^3$, which is about 6 electron volts.

Now the number of constrictions per unit length of dislocation line in thermal equilibrium is about $(1/a)\exp(-W/\boldsymbol{k}T)$. Taking $a = 10^{-8}$ cm and $\boldsymbol{k}T = \frac{1}{40}$ ev (corresponding to room temperature), we obtain only one constriction in 10^{90} cm length of dislocation line, so that thermal fluctuations alone are not sufficient to form any significant number of constrictions at room temperature. Even at 600°c we find the average distance between constrictions is about 10^{17} cm.

The energy W of the constriction may be reduced, and hence the number of constrictions formed increased, by the action of an applied stress; for we have seen that W depends on the stress through the width d of the dislocation. For a fluctuating stress d must be found from equation (4). In the case of a group of n dislocations piled up against a barrier we may use the results of Eshelby *et al.* to obtain the variation in the stress. Then we find that the width of the νth dislocation from the obstacle ($\nu \ll n$) will be about

$$d = \tfrac{1}{4}(a\nu/n)(\pi G/\sigma)^{1/2} \qquad \ldots\ldots(13)$$

and, with any reasonable values of the constants, this is small compared with the width of dislocation under zero stress. Consequently the energy which is given by (12) and (13) will be greatly reduced, so that constrictions will readily be produced in dislocations at the head of a piled-up group.

§3. Locked Dislocations

Lomer (1951) and Cottrell (1952) have pointed out that two extended dislocations in different glide planes may react on meeting to form a complex unable to glide. If the dislocations are $\frac{1}{2}a[10\bar{1}]$ in (111) and $\frac{1}{2}a[011]$ in $(11\bar{1})$, they will meet in the line $[1\bar{1}0]$ and, on reacting, give the three imperfect

dislocations with Burgers vectors $\mathbf{b}_P = \frac{1}{6}a[110]$, $\mathbf{b}_Q = \frac{1}{6}a[11\bar{2}]$ and $\mathbf{b}_R = \frac{1}{6}a[112]$ (figure 2). Such complexes are to be expected in the piled-up groups which occur in a work-hardened metal, and prevent the group from dispersing when the applied stress is reduced (Mott 1952).

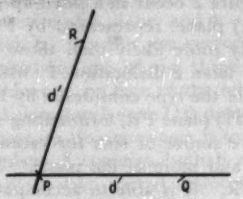


Figure 2.

To estimate the energy of a constriction in the complex we must first find the separation of the dislocations. We assume that the equilibrium position is symmetrical so that we have only one parameter $d' = PQ = PR$ to determine. As the dislocations are all parallel to the line $[1\bar{1}0]$, they are of pure edge type. Now Nabarro (1952a) has shown that the energy per unit length of interaction of two edge dislocations $\mathbf{b}'$ and $\mathbf{b}''$, the shortest distance between which is a vector $\mathbf{r}$, is

$$-\frac{G}{2\pi(1-\nu)}\left\{\mathbf{b}'\mathbf{b}''\log r - \frac{(\mathbf{b}'\mathbf{r})(\mathbf{b}''\mathbf{r})}{r^2}\right\}.$$

Summing contributions of this form for three pairs of dislocations, we find that in our case the energy of interaction is

$$-\{Ga^2/36\pi(1-\nu)\}\log d' + \text{terms independent of } d'.$$

The energy due to the stacking fault is $2F_0 d'$. F_0 may be expressed in terms of the distance d between the two partials $\mathbf{b}_1 = \frac{1}{6}a[1\bar{2}1]$ and $\mathbf{b}_2 = \frac{1}{6}a[2\bar{1}\bar{1}]$ of a free extended dislocation. We have

$$F_0 = \frac{G\mathbf{b}_1\mathbf{b}_2}{2\pi(1-\nu)d} = \frac{Ga^2}{24\pi(1-\nu)d}.$$

Hence that part of the energy of the complex which depends on d' is

$$-\frac{Ga^2}{36\pi(1-\nu)}\log d' + \frac{Ga^2}{12\pi(1-\nu)d}d'$$

and this is a minimum when $d' = \frac{1}{3}d$, giving the equilibrium spacing of the dislocations in the complex.

We may form a constriction by uniting the dislocations P and Q at a point. Since P must lie along the intersection of two glide planes, it will remain straight and undeformed, so that the constriction will not be symmetrical. The calculations of § 2, however, are easily modified to suit the present case. We then obtain for the energy

$$W = 1{\cdot}7(A'T)^{1/2}d' \qquad \ldots\ldots(14)$$

where

$$A' = G\mathbf{b}_P\mathbf{b}_Q/2\pi = Ga^2/36\pi$$

and T is, as before, given by (2). On substituting these values (14) becomes

$$W = 6\times 10^{-3}\, Ga^2 d\,\{\log(R/r_0)\}^{1/2}. \qquad \ldots\ldots(15)$$

On comparing (12) and (15) we see that the energy required to form a constriction in a locked dislocation is only 0·35 of that required in a free dislocation. Hence there will be, in similar conditions, many more jogs on a locked dislocation than on a free one, so that the locked dislocation may be expected to climb the faster.

If the dislocations of figure 2 occur in a piled-up group, the other members of which all lie in the (111) plane, represented by PQ in the figure, then jogs uniting P and Q are clearly more likely than those uniting P and R. When P and Q are combined we have a dislocation P′ with Burgers vector $\frac{1}{3}a[11\bar{1}]$. This is a sessile dislocation of the type considered by Frank (1949); it represents a close-packed plane, the $(11\bar{1})$ plane PR, terminating at the dislocation. Such a dislocation is a very effective source or sink for vacancies, as vacancies arriving there simply remove more and more of the plane. That is, the dislocation P′ in climbing moves along PR. R is able to accompany it by glide, so that the complex as a whole can climb out of the piled-up group. Once the locked dislocations have been removed from the piled-up group in this way, the free dislocations can disperse by glide.

§4. Jogs

A jog is a point at which a dislocation line is displaced from one atomic plane to another. A jog will be formed in a dislocation **b** when it crosses a screw dislocation $\mathbf{b}_s$ which intersects the glide plane. Figure 3 (*a*) represents a jog in a perfect dislocation; here $QR = \mathbf{b}_s$ (in magnitude and direction) and PQ and RS

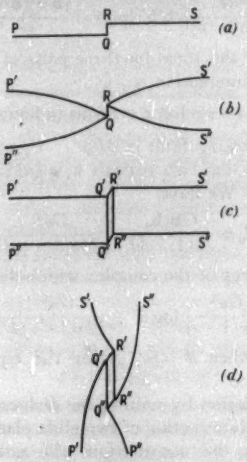


Figure 3.

lie in adjacent atomic planes. If the dislocation is of a type which can dissociate it will certainly dissociate near the points P and S; we must now consider whether it will dissociate at Q and R.

It will be convenient to use the notation of Thompson (1953). A tetrahedron whose faces are parallel to the {111} planes is used as a basis of

reference. Vertices of the tetrahedron are denoted by A, B, C, D, the mid-points of the opposite faces by α, β, γ, δ, respectively, and faces on which they lie by a, b, c, d. The edges of the tetrahedron **AB**, **BC**, etc., represent the Burgers vectors of all possible perfect dislocations, and these may dissociate into partial dislocations with Burgers vectors **A**γ, γ**B**, etc.

We take a to be the operative slip plane, and $\mathbf{b}=\mathbf{BC}$ is the Burgers vector of a dislocation in it. Then $\mathbf{b}_s$, the Burgers vector of an intersecting screw dislocation, will be represented by one of the three edges of the tetrahedron meeting at A. If $\mathbf{b}_s=\mathbf{AD}$, there is no face of the tetrahedron containing both **b** and $\mathbf{b}_s$, so that the short dislocation length QR does not have any {111} plane as its glide plane. In this case we do not expect it to dissociate (figure 3(*b*)).

If $\mathbf{b}_s=\mathbf{AB}$ or **AC**, the plane d contains both **b** and $\mathbf{b}_s$ and QR is able to dissociate in this plane into the two partial dislocations **B**δ and δ**C**. If the dislocation **b** is an edge the dissociation of QR leads to an extended dislocation of uniform width (figure 3(*c*)). Where the ribbon of stacking fault bends, at Q′Q″ and R′R″, we must have one of Thompson's stair-rod dislocations to obtain conservation of Burgers vector. The jogs of figures 3(*b*) and 3(*c*) may be distinguished as 'contracted' or 'extended' jogs respectively; this is in accord with Thompson's description of dislocation nodes. Now let the dislocation **b** rotate from the edge position toward the screw, when the step Q′Q″R″R′ makes an increasing oblique angle with the dislocation line. Figure 3(*c*) shows the configuration close to the screw position. From the figure it is clear that the line tension of the dislocations will cause the dislocation S″R″ to move its point of attachment along R″R′ to R′, and similarly Q′P′ will join up at Q″. Thus we now obtain a contracted jog. However, QR still possesses its ability to dissociate in the plane d, though when Q and R are only one atomic space apart dissociation is hardly likely to occur. However, if Q and R move apart, and so lengthen QR, a process involving only glide, dissociation may be expected. This will result in the dislocation as a whole changing its plane of dissociation from a to d. Mott (1952) has given an explanation of cross slip which depends on the ability of a screw to change its plane of dissociation; we see that the occurrence of a jog will provide a simple means for it to do so.

We may also note that jogs in a screw dislocation are free to move along the dislocation line. If there are several jogs in a dislocation it will be energetically advantageous for the jogs to move together and unite. This is clearly the case for jogs of opposite sign, for these will then cancel one another. It will also be true for jogs of the same sign, for the number of points at which the dislocation is contracted will then be reduced. The dislocation of figure 4(*a*), which has two jogs, is contracted at the four points P, Q, R and S. When the two jogs unite to give a single jog of double size, as in figure 4(*b*), the dislocation is contracted at only two points, U and V.

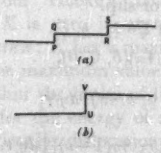


Figure 4.

The energy of a contracted jog will differ from that of a constriction by approximately the energy required to form the short length of dislocation QR (figure 3 (*b*)). The elastic energy of such a short dislocation will be small, and we expect most of the energy to arise from the atomic misfit in the core of the dislocation. Now Nabarro (1947) has shown that for a long straight dislocation the misfit energy is $Gb^2/4\pi$ per unit length. Assuming this also holds in our case, we obtain for the energy of QR $Ga^3/8\sqrt{2}\pi$. Since this is small compared with the energy of the constriction it may be neglected, and the energy of the jog will then be given by equation (12).

The energy of an extended jog will be smaller, but since it depends almost entirely on effects near the core of the dislocation it is difficult to make any reliable estimate of it. However, if, as seems possible, an extended jog goes through a contracted stage in the course of its formation, equation (12) will also give the activation energy of the extended jog.

ACKNOWLEDGMENT

I wish to thank Professor N. F. Mott for suggesting the problem, and Dr. F. C. Frank for helpful discussions.

APPENDIX

THE LINE TENSION OF A DISLOCATION

A derivation of the expression for the line tension of a dislocation is given here which is not subject to the assumption that the dislocation line is nearly straight.

Burgers (1939) has shown that the displacement $\mathbf{u}$ at a point $\mathbf{x}$ due to a dislocation with Burgers vector $\mathbf{b}$ can be written in the form

$$\mathbf{u}=\mathbf{u}^*+\mathbf{u}^{**}+\text{grad}\,\psi. \qquad \text{.....(A 1)}$$

Here $\mathbf{u}^*=\Omega\mathbf{b}/4\pi$,

where Ω is the solid angle subtended by the dislocation line at x;

$$\mathbf{u}^{**}=\frac{\mathbf{b}}{4\pi}\wedge\int\frac{d\boldsymbol{\xi}}{r}$$

where $d\boldsymbol{\xi}$ is an element of the line of the dislocation, $r=|\mathbf{x}-\boldsymbol{\xi}|$, and the integral is taken once round the dislocation loop; and

$$\psi=\frac{1}{8\pi(1-\nu)}\int\mathbf{b}\wedge(\mathbf{x}-\boldsymbol{\xi})\frac{d\boldsymbol{\xi}}{r}$$

where ν is Poisson's ratio.

We shall consider only dislocation loops which lie wholly in the glide plane. Our treatment follows closely that of Nabarro (1952 b) for a circular dislocation. We take the origin in the glide plane, the z axis normal to the glide plane, and the x axis parallel to the direction of slip.

Then $\mathbf{b}=(b, 0, 0)$. We put $\boldsymbol{\xi}=(\xi, \eta, \zeta)$.

Then
$$\left.\begin{aligned}\mathbf{u}^*&=(\Omega/4\pi)(b, 0, 0),\\ \mathbf{u}^{**}&=(0, 0, b/4\pi)\int d\eta/r,\\ \psi&=-\{bz/8\pi(1-\nu)\}\int d\eta/r.\end{aligned}\right\} \qquad \text{.....(A 2)}$$

We see that **u**, and hence all the stresses, can be expressed in terms of Ω and $\int d\eta/r$.

The elastic energy of the dislocation is

$$W = -\tfrac{1}{2}b\iint p_{zz}\,dx\,dy, \qquad \ldots\ldots(A\,3)$$

where the integral is taken over the area enclosed by the loop. p_{zz} is the shear stress $G(\partial u_x/\partial x + \partial u_z/\partial z)$, and G is the rigidity modulus. Then from (A 1), (A 2) and (A 3)

$$W = -\frac{Gb^2}{8\pi}\iint\left(\frac{\partial\Omega}{\partial z}\right)_{z=0} dx\,dy + \frac{G\nu b^2}{8\pi(1-\nu)}\iint dx\,dy\,\frac{\partial}{\partial x}\int\frac{d\eta}{r}. \qquad \ldots\ldots(A\,4)$$

Now Ω is equal to the scalar magnetic potential of a unit current flowing along the line of the dislocation. The integral involving Ω is equal to the mutual inductance of two circuits, separated by a short distance, and whose geometrical form is that of the line of the dislocation. This mutual inductance may be expressed as the line integral of the vector potential, so that (A 4) may be written

$$W = \frac{Gb^2}{8\pi}\left\{\int d\mathbf{s}\int\frac{d\boldsymbol{\sigma}}{r} + \frac{\nu}{1-\nu}\int dy\int\frac{d\eta}{r}\right\}$$

where $d\mathbf{s}$ and $d\boldsymbol{\sigma}$ are elements of the line of the dislocation, $d\mathbf{s}$ depending on $\mathbf{x}$ and $d\boldsymbol{\sigma}$ on $\boldsymbol{\xi}$. Since

$$d\mathbf{s}\,.\,d\boldsymbol{\sigma} = \left(\frac{dx}{ds}\frac{d\xi}{d\sigma} + \frac{dy}{ds}\frac{d\eta}{d\sigma}\right)ds\,d\sigma$$

we have

$$W = \frac{Gb^2}{8\pi}\int ds\int\left\{\frac{dx}{ds}\frac{d\xi}{d\sigma} + \frac{1}{1-\nu}\frac{dy}{ds}\frac{d\eta}{d\sigma}\right\}\frac{d\sigma}{r}, \qquad \ldots\ldots(A\,5)$$

so that the dislocation has energy T per unit length, given by

$$T = \frac{Gb^2}{8\pi}\left\{\frac{dx}{ds}\int\frac{d\xi}{r} + \frac{1}{1-\nu}\frac{dy}{ds}\int\frac{d\eta}{r}\right\}. \qquad \ldots\ldots(A\,6)$$

We may call T the line tension of the dislocation. Its value will in general vary from point to point along the dislocation line.

If P(x, y) is any point on the dislocation line, we take an origin at P, and axes t, n along the tangent and normal at P to the dislocation line; ξ and η are then to be expressed in terms of t and n. Further, we define θ as the angle between the Burgers vector and an element of the dislocation line at P, so that $dx/ds = \cos\theta$ and $dy/ds = \sin\theta$. Then the line tension at the point P becomes

$$T = \frac{Gb^2}{8\pi}\left(\frac{1-\nu\cos^2\theta}{1-\nu}\right)\int\left\{\frac{1}{r} - K\frac{dn}{dt}\frac{1}{r}\right\}dt, \qquad \ldots\ldots(A\,7)$$

where $r^2 = t^2 + n^2$. Since the integral of $1/r$ diverges when r tends to 0, we must as usual take $r > r_0$, where r_0 is the distance from the dislocation within which departures from Hooke's law become important. Also $K = \nu\sin 2\theta/\{2(1-\nu\cos^2\theta)\}$. K is zero if, at the point P, the dislocation is either purely edge or purely screw; K has a maximum value of $\nu/(1-\nu)^{1/2}$ when $\theta = \frac{1}{2}\cos^{-1}\nu/(2-\nu)$. If $\nu = \frac{1}{3}$ the maximum value of K is $1/\sqrt{6}$.

Equation (A 7) is exact within the limits of linear elastic theory. It clearly leads to the usual expression for the energy of a straight dislocation, and it is also easy to verify that for a circular dislocation it gives the value of the energy obtained by Nabarro (1952 b). Apart from these simple cases, the exact

evaluation of the energy from (A 7) appears to involve awkward integrations. We may, however, obtain a useful approximation to T as follows.

We note that dn/dt is zero at the point P, and provided the radius of curvature is large compared with r_0, dn/dt will remain small until r is large, so that the second term of the integrand is always small. Further, unless P is a point of inflection, dn/dt has opposite signs at either side of P, so that the contributions to the integral tend to cancel one another out. Hence the value of the integral in (A 7) is determined mainly by the first term of the integrand. The form of this term suggests that the integral can be written as $2 \log (R/r_0)$, the factor 2 occurring as there are two branches of the dislocation line extending from P. The precise value of R will depend on the point P, but provided the dislocation line is not of too irregular a form, R is unlikely to vary in order of magnitude, so that the logarithm is approximately constant. We expect R to be of the order of the dimensions of the dislocation loop.

Also the factor $(1-\nu \cos^2 \theta)/(1-\nu)$ varies only between 1 and $1/(1-\nu)$, so that we shall not be greatly in error if we approximate it by unity (or by any other constant within these limits). With these approximations the line tension is constant and given by

$$T = \frac{Gb^2}{4\pi} \log \frac{R}{r_0}. \qquad \ldots\ldots\ldots (A\,8)$$

This may be compared with the expression for the line tension obtained by Mott and Nabarro (1948).

References

Burgers, J. M., 1939, *Proc. Acad. Sci. Amst.*, **42**, 293.

Cottrell, A. H., 1952, *Phil. Mag.*, **43**, 645.

Eshelby, J. D., Frank, F. C., and Nabarro, F. R. N., 1951, *Phil. Mag.*, **42**, 351.

Frank, F. C., 1949, *Proc. Phys. Soc.* A, **62**, 202.

Heidenreich, R. D., and Shockley, W., 1948, *Bristol Conference on the Strength of Solids* (London : Physical Society), p. 57.

Lomer, W. M., 1951, *Phil. Mag.*, **42**, 1327.

Mott, N. F., 1952, *Phil. Mag.*, **43**, 1151.

Mott, N. F., and Nabarro, F. R. N., 1948, *Bristol Conference on the Strength of Solids* (London : Physical Society), p. 1.

Nabarro, F. R. N., 1947, *Proc. Phys. Soc.*, **59**, 256 ; 1952 a, *Advances in Physics*, **1**, 309 ; 1952 b, *Ibid.*, 334.

Thompson, N., 1953, *Proc. Phys. Soc.* B, **66**, 481.

RESEARCH NOTES

Negative Ion Formation in CCl_4 and $TiCl_4$

By J. MARRIOTT, R. THORBURN and J. D. CRAGGS
Department of Electrical Engineering, University of Liverpool

MS. received 28*th December* 1953 *and in amended form* 10*th February* 1954

The purpose of this note is to give new data, obtained with a Lozier ionization apparatus, relating to negative ion formation in CCl_4 and $TiCl_4$. Previous studies with this type of apparatus on O_2, Br_2 and certain other molecules are described by Massey (1950), who also (p. 59) shows how the kinetic energy of the fragments may be measured (see figure 2).

The reactions studied in the present case are :

$$XY + e \rightarrow X + Y^- \qquad \ldots\ldots(1)$$

and

$$XY + e \rightarrow X^+ + Y^- + e. \qquad \ldots\ldots(2)$$

The appearance potentials for X^+ and Y^-, i.e. $A(X^+)$ and $A(Y^-)$, were measured, together with the kinetic energy of the Y^- and X^+ ions, in a Lozier apparatus (Lozier 1934), following a necessary preliminary investigation in a mass spectrometer to establish that a single negative or positive ion dominated

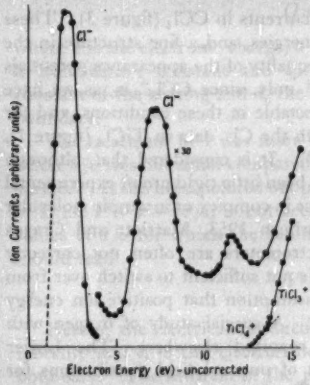


Figure 1. Predominating negative and positive ions in $TiCl_4$. Current units are different for ions of different sign.

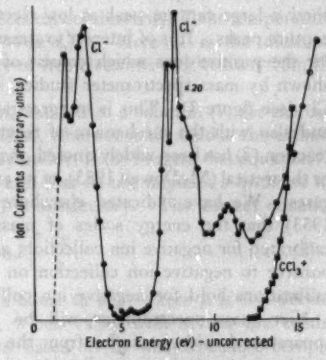


Figure 3. Predominating positive and negative ions in CCl_4. Current units are different for ions of different sign.

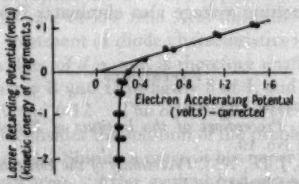


Figure 2. Lozier plot for Cl^- ions in $TiCl_4$.

each respective dissociation pattern. Unless this can be proved initially, the Lozier apparatus does not give easily interpretable results.

The apparatus was first checked with O_2, in which O^- is the predominating negative ion at the pressures used ($\sim 10^{-5}$ mm Hg). The measurement of initial kinetic energy of the O^- (reaction (1), above) together with the dissociation energy of O_2, taken as 5·1 ev, enabled the electron affinity of O^- to be estimated as $2{\cdot}0 \pm 0{\cdot}2$ ev, in agreement with Lozier's original value of $2{\cdot}2 \pm 0{\cdot}2$ ev. The appearance potential standard was $A(O_2^+) = 12{\cdot}2$ ev.

Figure 1 shows the variation of Cl^- current with electron energy in $TiCl_4$. Also in figure 1 are plotted the positive ion currents in $TiCl_4$; again, previous mass spectrometer work had shown that $TiCl_4^+$ appeared at $11{\cdot}7 \pm 0{\cdot}2$ ev and that $TiCl_3^+$ appeared at $12{\cdot}88 \pm 0{\cdot}05$ ev, so that the break in the curve of figure 1 is due to the superposition of the $TiCl_4^+$ and $TiCl_3^+$ currents. Figure 1 shows that $TiCl_3^+$ appears near the potential at which a sudden increase in Cl^- occurs; this is interpreted as the threshold for reaction (2), above. Typical data taken with the Lozier apparatus show $A(TiCl_3^+) = 12{\cdot}9 \pm 0{\cdot}5$ ev and $A(Cl^-) = 12{\cdot}6 \pm 0{\cdot}3$ ev, taking $A(TiCl_4^+) = 11{\cdot}7$ ev as standard. Kinetic energy data (figure 2) show that the appearance potential for the first Cl^- peak of figure 1 occurs, for extrapolation to zero initial kinetic energy, at $0 \pm 0{\cdot}1$ ev, which is to be expected if the reaction products are unexcited and if the electron affinity of Cl, taken as 3·8 ev, is greater than the $TiCl_3$–Cl bond dissociation energy. Apparently the latter quantity has not been determined.

Finally, we report measurements of Cl^- currents in CCl_4 (figure 3). These show a large capture peak at low electron energies, and a fine structure in the capture peaks. It is of interest to stress the equality of the appearance potentials for the positive ions which consist of CCl_3^+ only, since CCl_4^+ is, as we have shown by mass spectrometer studies, undetectable in these conditions, and for Cl^- (see figure 3). This is in agreement with the Cl^- data in $TiCl_4$ (figure 1), and also with the mechanism of reaction (2). It is considered that, although reaction (2) has been widely quoted, there has been little rigid proof, experimental or theoretical (McDowell 1953), of its existence in complex or in simple molecular gases. We have indicated elsewhere (Thorburn 1953, Marriott and Craggs 1953) that the energy scales of mass spectrometers are often not correctly calibrated for negative ion collection, and it is not sufficient to switch over from positive to negative ion collection on the assumption that positive ion energy calibrations hold for negative ion collection. A special study of oxygen with a Nier mass spectrometer will be fully reported elsewhere. The Lozier apparatus should not suffer from the defect of unequal field distributions for positive and negative ion collection, since in the Lozier apparatus the shielding of the filament region from the ionization region is more complete than in a normal mass spectrometer. Further, in the Lozier apparatus the absence of the high tension accelerating voltage also eliminates field penetration into the ionization region.

References

Lozier, W. W., 1934, *Phys. Rev.*, **46**, 268.

McDowell, C. A., 1953, *Proceedings of the Institute of Petroleum Conference on Mass Spectrometry*, Paper No. 10.

Marriott, J., and Craggs, J. D., 1953, *Proceedings of the Institute of Petroleum Conference on Mass Spectrometry*, Paper No. 13.

Massey, H. S. W., 1950, *Negative Ions* (Cambridge : University Press).

Oxide Cathodes in Accelerating Fields

By P. N. DAYKIN and D. A. WRIGHT

Communication from the Research Staff of the M.-O. Valve Company at the G.E.C. Research Laboratories, Wembley, England

MS. received 7th January 1954

In a recent paper (Wright and Woods 1952, to be referred to as I) experimental and theoretical work was described dealing with the effect of the potential gradient at the surface of an oxide coating on its emission. The oxide surface was supposed to consist of a homogeneous dielectric, i.e. no effects of porosity were included. Moreover, the ionized donor centres were supposed not to move under the influence of the fields in the coating. A more elaborate treatment would be necessary to include these effects.

A theoretical treatment was first given by Morgulis (1947), dealing with the case where current density was small, and where the field at the cathode was large. The limitation with regard to current density J was given by

$$Q = \boldsymbol{e} J l_0 / \sigma \boldsymbol{k} T \ll 1,$$

where $\boldsymbol{e}$ is the electronic charge, σ the electrical conductivity, T the temperature, $\boldsymbol{k}$ Boltzmann's constant and $l_0 = (K\boldsymbol{k}T/4\pi\boldsymbol{e}^2 n_0)^{1/2}$. Here K is the dielectric constant of the coating and n_0 the density of free electrons in it remote from the boundary.

The treatment in I was more general, and led to a solution (7) for the case where J/σ was not small compared with X/K, where X is the field at the cathode surface. It has been pointed out privately by Mr. R. E. Burgess that equation (2) in I, i.e. $V' = V - \boldsymbol{e}J/\sigma$, is not accurate, and that its use imposes a limit on the values of J/σ for which the solution (7) is satisfactory. The correct equation should read

$$V' = V - \boldsymbol{e} J \int_0^x \frac{dx}{\sigma}. \qquad \ldots\ldots (2)'$$

Here σ, the electrical conductivity, is equal to $\boldsymbol{e}n\mu$, where μ is the electron mobility and n is the electron density, which varies with x near the coating boundary. The solution of the problem based on (2)′ would be more complex than that given in I. Using equation I (2), the solution I (7) is satisfactory for values of J limited as in Morgulis' case, i.e. $Q \ll 1$. Provided this is complied with, equations (7) and (8) are more widely applicable than Morgulis' solution, in that they include the case of low field X at the cathode, including zero field.

In the subsequent treatment of diode characteristics in I, most of the analysis is restricted to small values of J/σ, and is therefore unaffected, but there will be some inaccuracy in curves B and D in figure 8 of I, and in two of the values for $\delta\chi_0$ in table 1, i.e. 0·29 and 0·12. The other values are not affected.

We were about to reconsider the solution of the problem, replacing (2) by (2)′, when we received from Dr. Shindo a copy of his paper on this subject (Shindo 1951) which we had not seen. After we had had some private discussion with Dr. Shindo, he wrote a further paper (Shindo 1953). We consider that his method

of solving this problem is equivalent to that involved in combining the equations (2)′ and I (3), but is to be preferred. We have therefore not proceeded with our solution. Shindo's treatment is general, and is not restricted to small current, but his solution is obtained by successive approximations, and the zero order solution is limited to the condition

$$-1 \ll g_l = \frac{\Delta E}{e n_0 l_0} \frac{K}{8\pi} \frac{Q+(Q^2+4)^{1/2}}{2} \ll 1.$$

ΔE is the difference between the field X/K immediately inside the coating and the field J/σ in the body of the coating.

It is perhaps worth recording that we have checked Shindo's zero order solution, solving his equations (2.9) and (2.10) by eliminating $\exp(-g_l)$, and we have also carried out the first and second order approximations to the solution of this equation (2.6). Shindo had compared the first order solution with the zero order in his 1951 paper, but did not do so for the complete solution in his 1953 paper. It is very desirable to do this because the zero order approximation applies for values of g_l small compared with unity, yet table 1 in the 1953 paper gives calculations of current for values of g_l as large as 0·386. The solution obtained below taking the second order approximation still requires $g_l < 1$, but not $g_l \ll 1$, and is in fact accurate at least to $g_l = 0{\cdot}7$. We find that the plot of log(current) against square root of field is not significantly different, comparing the zero and the second order approximations, for values of g_l up to 0·4. The solution of (2·6) has been carried out by the following procedure: We write

$$\exp\{-g(x)\} = 1 - f(x), \qquad \ldots\ldots(1)$$

and obtain for Shindo's equation (2.6)

$$f'' - Qf' - f = ff'' - (f')^2 - \tfrac{3}{2}f^2 + \tfrac{1}{2}f^3. \qquad \ldots\ldots(2.6)'$$

The solution, which is asymptotic to $f_0 \exp\{\alpha_1(x'-L)\}$ for small f, can be expressed as the power series

$$f(x) = \sum_{n=1}^{\infty} f_{n-1} \exp\{n\alpha_1(x'-L)\}. \qquad \ldots\ldots(2)$$

Here $L = d/l_0$, d = coating thickness, $x' = x/l_0$. When this is substituted in (2.6)′ the recurrence relation between the coefficients f_n is found to be

$$[n(n-1)Q\alpha_1 + (n^2-1)]f_{n-1} = \sum_{l=1}^{n-1}[(n-l)(n-2l)\alpha_1^2 - \tfrac{3}{2}]f_{n-l-1}f_{l-1} + \tfrac{1}{2}\sum_{l=1}^{n-1} f_{n-l-1}\sum_{k=1}^{l-1} f_{l-k-1}f_{k-1},$$

with the help of the equation $\alpha_1^2 - Q\alpha_1 - 1 = 0$, and the sum rule

$$\sum_{n=1}^{\infty}\sum_{l=1}^{\infty} a_n a_l y^{n+l} = \sum_{n=1}^{\infty} y^n \sum_{l=1}^{n-1} a_{n-l}a_l.$$

The first two coefficients are

$$f_1 = \frac{-\frac{3}{2}f_0^2}{2Q\alpha_1+3}, \quad f_2 = \frac{\frac{1}{4}(9-Q\alpha_1)f_0^3}{(6Q\alpha_1+8)(2Q\alpha_1+3)}. \qquad \ldots\ldots(3)$$

The integral in Shindo's equation (2.3) can be expressed in terms of this solution to the required order of approximation. To second order this is

$$\int_0^l \sinh g(x)\,dx = \frac{l_0}{\alpha_1}\{g_l - \tfrac{1}{2}f_1 - \tfrac{1}{4}f_0^2 - \tfrac{2}{3}f_0 f_1 - \tfrac{1}{8}f_0^3 - \tfrac{1}{2}f_0^2 f_1\}$$

where g_l has been re-introduced by the expansion

$$g_l = -\log(1-f(l)) \simeq f + \tfrac{1}{2}f^2 + \tfrac{1}{3}f^3.$$

The term in f_2 is negligible for the values of g_l considered and has been dropped, but the largest third order term is not small and has been retained. Shindo's equation (3.1) can then be written as

$$E = K\left\{\exp(-g_l)\frac{\Phi}{\mu} - \frac{\beta g_l}{\alpha_1}\right\} + \frac{K\beta}{\alpha_1}(\tfrac{1}{2}f_1 + \tfrac{1}{4}f_0^2 + \tfrac{2}{3}f_0f_1 + \tfrac{1}{6}f_0^3 + \tfrac{1}{2}f_0^2f_1). \quad \ldots\ldots(3.1)'$$

By using Shindo's equation (3.2) to eliminate g_l we obtain the equation relating field strength and current:

$$E = K\left\{\frac{J}{en_0\mu} + \frac{\beta}{\alpha_1}\log\left(\frac{J}{en_0\Phi_t}\right)\right\} - \frac{K\beta}{\alpha_1}\frac{4{\cdot}39}{T}\sqrt{E}$$
$$+ \frac{K\beta}{\alpha_1}\{\tfrac{1}{2}f_1 + \tfrac{1}{4}f_0^2 + \tfrac{2}{3}f_0f_1 + \tfrac{1}{6}f_0^3 + \tfrac{1}{2}f_0^2f_1\}, \quad \ldots\ldots(4)$$

where

$$\exp(-g_l) = \frac{J}{en_0\Phi_0}\exp(-4{\cdot}39E^{1/2}/T) = 1 - f_0 - f_1 - f_2. \quad \ldots\ldots(5)$$

The correction terms in (4) were first neglected and the field strength corresponding to a given value of J was found by trial. Using these values of E and J and the values of f_1 and f_2 from (3), the cubic equation (5) was solved for f_0 by trial. The correction terms in (4) were computed and the new value of E found again by trial. Using Shindo's values for the constants n_0, μ and K, the corrections to the 800°K curve were found to be negligible for values of g_l ranging from 0·39 to −0·49. For the higher temperatures the corresponding currents are larger and the coefficients f_1 and f_2 consequently even smaller. The smallness of the corrections is explained by the partial cancellation of the first order terms, and this occurs because f_1 is negative.

Acknowledgments

The authors desire to tender their acknowledgments to the M.-O. Valve Co. Ltd., on whose behalf the work described in this publication was carried out.

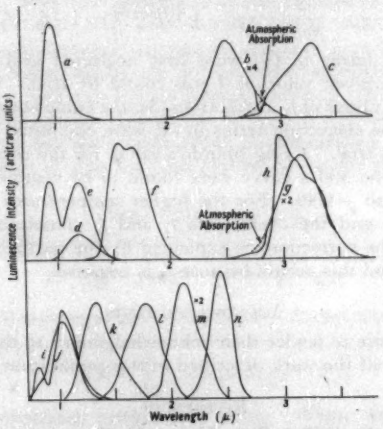
References

Morgulis, N., 1947, *J. Phys., Moscow*, **11**, 67.
Shindo, T., 1951, *J. Phys. Soc. Japan*, **6**, 352; 1953, *Ibid.*, **8**, 494.
Wright, D. A., and Woods, J., 1952, *Proc. Phys. Soc.* B, **65**, 134.

LETTERS TO THE EDITOR

Phosphors Emitting Infra-Red Radiation

In recent investigations we have made successful attempts to produce phosphors with emission spectra lying in the infra-red region beyond 1 μ. The spectra of these phosphors, shown in the accompanying figure, were analysed with a double-prism spectrometer and a lead sulphide or cooled lead telluride cell as detector, in conjunction with a 'chopper disc' light beam modulator, 800 c/s tuned amplifier and pen recorder. Either fused quartz or lithium fluoride prisms were used. Specimens were excited at the respective wavelength for optimum emission by a tungsten lamp provided with suitable filters.



Spectra of infra-red emitting phosphors. C denotes spectra corrected to equal energy conditions. PbS/Q denotes spectra recorded by PbS cell through quartz prisms, PbTe/LiF denotes spectra recorded by PbTe/cell through LiF prisms, PbS/LiF denotes spectra recorded by PbS cell through LiF prisms.

a, Cu_2O, 290°K, C ; *b*, PbS, 290°K, PbTe/LiF ; *c*, PbS, 90°K, PbTe/LiF ; *d*, CdSe–Cu, 290°K, PbS/Q; *e*, CdSe–Cu, 90°K, PbS/Q; *f*, ZnS–Cu, 290°K, PbS/LiF; *g*, ZnS–Co, 290°K, PbTe/LiF; *h*, ZnS–Co, 90°K, PbTe/LiF; *i*, CdS–Cu, 290°K, C; *j*, CdS–Cu, 90°K, C; *k*, CdS (70%) HgS (30%)–Cu, 290°K, C; *l*, Cd S(25%) HgS (75%)–Cu, 290°K, PbS/Q; *m*, HgS, 290°K, PbS/Q; *n*, HgS, 90°K, (PbS/Q.

By extrapolation from well-known results for zinc–cadmium sulphide phosphors we expect the emission of copper activated cadmium sulphide to be in the 1 μ region. This was confirmed by Grillot and Guintini in 1953. Spectra for our own specimens of this phosphor are given in curves *i* and *j* for 290°K and 90°K respectively. Extending investigations to heavier members of this phosphor group we have produced cadmium selenide activated by copper, the spectra for 290°K and 90°K being given in curves *d* and *e* respectively. For

mixed cadmium–mercury sulphides the emission moves to longer wavelengths with increasing mercury content (curves *k*, *l*, *m* and *n*) so that for mercury sulphide the room temperature emission has a peak at 2·1 μ. On cooling the specimen to 90°K the peak moves to 2·45 μ. Two other infra-red emitting phosphors occur in this group. The first is copper activated zinc sulphide which, when excited by near infra-red radiation at low temperature, gives an emission with maxima at 1·5 and 1·65 μ (curve *f*). Specimens prepared under particular conditions show this emission at room temperature. The second phosphor is cobalt activated zinc sulphide which, when excited by radiation in the 2 μ region, shows emission at 3·0–3·3 μ. The spectrum does not shift with temperature, but on cooling the phosphor it exhibits a fine structure. The latter is considerably obscured by the necessity of using wide spectrometer slits to record the spectrum. The luminescence transitions thus seem to be characteristic of those within the cobalt ions. Further investigations of these two phosphors are being made in view of the importance of copper and cobalt as activators in more conventional zinc sulphide phosphors (Hoogenstraten and Klasens 1953).

Some further infra-red emitting phosphors have been investigated following reports by Russian workers. Lashkarev and Kossonogova (1946) reported a 1 μ emission from cuprous oxide layers prepared as metal rectifiers. The spectrum for cuprous oxide is given in curve *a*. The emission is, however, most intense when the specimen is copper-rich. The excitation spectrum follows the spectral response curve for the photoconductivity of cuprous oxide. Khlebnikov and Melamid (1948) have reported emission from caesium–antimony surfaces in photocells excited by near infra-red radiation. We have not so far succeeded in observing this emission. More recently Galkin and Korolev (1953) have observed luminescence in the 2·6 μ region from lead sulphide photoconductive cells. We have confirmed their observations and have made, by various methods, specimens of lead sulphide which show the spectra given in curves *b* and *c* for 290°K and 90°K respectively. The spectral shift with temperature corresponds to the shift of the photoconductive response limit with temperature, and the emission efficiency shows a similar dependence to that of photoconduction on the presence of oxygen (Moss 1949, Sosnowski, Starkiewicz and Simpson 1947).

It should be pointed out that great care is necessary in such experiments to avoid confusion of luminescence emission with black-body radiation from the specimen due to absorption and thermal dissipation of the exciting radiation. We hope to publish more detailed data later on other luminescence characteristics such as excitation, stimulation and quenching spectra and occurrence of thermal glow and phosphorescence.

Department of Physics,
University of Birmingham.
19th February 1954.

G. F. J. Garlick.
M. J. Dumbleton.

Galkin, L. N., and Korolev, N. V., 1953, *Dokl. Akad. Nauk., S.S.S.R.*, **92**, 529.
Grillot, E., and Guintini, P., 1953, *C. R. Acad. Sci., Paris*, **236**, 802.
Hoogenstraten, W., and Klasens, H. A., 1953, *J. Electrochem. Soc.*, **100**, 366.
Khlebnikov, V. E., and Melamid, A. E., 1948, *Dokl. Akad. Nauk., S.S.S.R*, **63**, 649.
Lashkarev, V. E., and Kossonogova, A. M., 1946, *Dokl. Acad. Sci., U.R.S.S*, **54**, 125.
Moss, T. S., 1949, *Proc. Phys. Soc.* B, **62**, 741.
Sosnowski, L., Starkiewicz, J., and Simpson, O., 1947, *Nature, Lond.*, **159**, 818.

REVIEWS OF BOOKS

Stress Waves in Solids, by H. KOLSKY. Pp. x+211. (Oxford: Clarendon Press, 1953.) 25*s*.

The theory of the propagation of transverse and longitudinal waves in elastic bodies dates from Navier, Poisson and Rayleigh in the last century but few results to test the theories were available at the time because of experimental difficulties and the fact that the nature of the coupling between different modes of vibration was not properly understood. The present text gives a useful survey of the basic theory of the subject and continues with a description of the experiments, starting with the well-known Hopkinson bar of 1914 and its modern refinement at the hands of R. M. Davies and others.

The second part of the book treats of waves in imperfectly elastic media—including the theory of visco-elastic bodies of the Maxwell and Voigt types—and the experimental work on the vibrations of such bodies both in the sonic and ultrasonic range of frequency. Finally plastic waves and shock waves are given a brief treatment.

The book is to be recommended as a concise summary of an important aspect of the solid state and the reader needing more information will find at the end a useful bibliography. E. G. R.

Mechanical Properties of Wood and Paper, by W. W. BARKAS, R. F. S. HEARMON and H. F. RANCE. Edited by R. MEREDITH. (Deformation and Flow: Monographs on the Rheology of Natural and Synthetic Substances.) Pp. xii+298. (Amsterdam: North-Holland Publishing Company, 1953.) 50*s*.

The mechanical properties of high polymers is a subject of considerable complexity. When these polymers are in the form of fibres, which are arranged anisotropically and further are able to incorporate varying quantities of water within their structure, the problem of interpreting the behaviour under applied stress becomes even more difficult. This book deals with two such substances and succeeds in making this apparently unpromising field one of interest to the physicist and the physical chemist as well as to the technologist.

The first section of the monograph has been written by Dr. W. W. Barkas and is entitled 'The Mechanical Properties of Wood and their Relation to Moisture'. Dr. Barkas has made many noteworthy contributions to this field and describes here how the elastic properties of wood may be correlated with its microscopic cellular structure. The relations between the rheological and thermodynamic aspects of the problem of swelling are discussed and it is shown how such relations may be derived without assuming any knowledge of the particular microscopic processes involved.

A chapter in this section by Mr. R. F. S. Hearmon discusses the elastic and plastic properties of wood. There is here a clear exposition of the elastic behaviour of this anisotropic material and the properties of plywood are related to the moduli of the constituent layers. The part of this chapter which deals with the plasticity of wood appeared rather less convincing, largely as a result of the empirical nature of this subject, and the absence of sufficient experimental data.

The second, and larger, section of the book has been written by Dr. H. F. Rance and is entitled 'The Mechanical Properties of Paper'. Dr. Rance says in his preface that this section of the book has a twofold aim, namely to interest the rheologist in paper-making and the 'paper-maker' in rheology. The first chapter gives a lucid account of the paper-making process and concludes with some comments on how the rheologist is involved in the problems of this industry and this theme is amplified in later chapters which discuss the interpretation of rheological behaviour and the practical importance of the measured rheological properties. The second chapter describes the conventional testing of paper and discusses the relation between the different mechanical tests; there are also chapters on the hygrosensitivity of paper, on the relation between the rheological behaviour of paper and that of other materials, and on the instrumentation of paper testing. Dr. Rance has a lively style and succeeds in showing that the paper-maker needs the rheologist and that the rheologist can find plenty to interest him in the field of paper-making if he can be persuaded to enter it.

The whole book is well produced and illustrated and has an adequate bibliography. H. KOLSKY.

Small Transformers and Inductors, by K. A. MACFADYEN. Pp. xii+237. (London: Chapman and Hall, 1953.) 37*s*. 6*d*.

There has long been the need for a book which treats the design of transformers for use at both power and communication frequencies on the basis of a common, systematic theory, and which so closely relates the design process to fundamental theory as to enable the non-specialist successfully to undertake the design of transformers for widely differing applications.

The author has fulfilled this need most admirably. In the compass of some 200 pages the reader is taken from the fundamentals of electromagnetism to the practical design of transformers of a variety of types, including power frequency, wide-band, high-Q and pulse transformers. The emphasis throughout is on a logical, and wherever possible synthetic, method of design in which processes of 'cut-and-try' are almost completely eliminated and the physical implication of the various steps may be clearly seen.

The first two chapters of the book give a compact presentation of the electric circuit and electromagnetic theory to be needed subsequently. Then follows the development of transformer equivalent circuits, taking as a starting point the 'perfect' transformer—one having finite winding inductances but perfect magnetic coupling—and introducing one by one the imperfections existing in practice. A novel feature is the representation of core losses by allowing the permeability to become complex; this imperfection is readily represented by slight modification of the equivalent circuit of the transformer. The effects of the various transformer imperfections on performance (e.g. frequency response characteristics, ferromagnetic distortion) are considered in detail.

Before going into the design of specific types of transformer, a very useful chapter is devoted to the calculation and control of the shunt and leakage inductance of windings, the effects of sectionalizing being discussed. Methods of inductance measurement are also described.

At this stage an adequate foundation has been laid, ready for proceeding with detailed transformer design to a wide range of practical requirements.

The rest of the book deals with the design of power transformers at 50 c/s and higher frequencies, wide-band transformers with and without d.c. polarization, instrument (voltage and current) transformers, high-Q transformers and chokes and pulse transformers.

The final chapter of the book gives valuable information on core materials, coil winding and insulating materials. With the aid of the graphs and tables provided there should be no difficulty in carrying the design work to a successful conclusion.

Written in a lucid and attractive style the book is easy to follow throughout. This must be attributed largely to the careful organization of the material, in particular of the theoretical development of the first five chapters, upon which the later design methods depend. The text is illustrated by many diagrams, graphs and worked examples, and frequent footnote references are given for further reading on the more specialized topics. A. R. B.

The Outer Layers of a Star, by R. v. d. R. Woolley and D. W. N. Stibbs. Pp. xi+306. (Oxford : Clarendon Press, 1953.) 40*s*.

Several books on the theory of stellar atmospheres have recently been published, but this admirable book still manages to fill a gap. The subject is a very extensive and rapidly expanding one, so that a summary of basic principles and review of progress are always welcome. The first chapters of the book are concerned with the theory of radiative equilibrium and line absorption. Here are considered Chandrasekhar's approximations to the solution of the equation of transfer, and exact solutions of the integral equations. All of this work is very necessary to a proper understanding of stellar atmospheres, but it might have been desirable to have laid more emphasis on the fact that the theories are built on the assumption of a plane stratified atmosphere in static equilibrium. The later chapters of the book deal with the outermost parts of stellar atmospheres, in particular the solar chromosphere and corona. Theories of these parts of the sun are continually changing and it would have been interesting to have had a more critical appreciation of them. Thus, there are more important criticisms of the accretion theory of the solar corona than those mentioned.

The book has no claims to being an exhaustive treatment of the subject, but its usefulness would have been increased if more references to the literature had been included. All these, however, are no more than minor criticisms of an excellent and strongly recommended book, which it has been a pleasure to read.

The printing, binding and arrangement of the text are of the high standard associated with the Oxford Clarendon Press. D. E. B.

Crystal Growth and Dislocations, by Ajit Ram Verma. Pp. xii + 182. (London : Butterworths Scientific Publications, 1953.) 30*s*.

The object of this book is to present an account of recent work in the field of crystal growth, and in particular that work bearing on the dislocation theory of crystal growth. The book is based on the author's Ph.D. thesis and some of its major limitations stem from that fact. One almost inevitable result is that much of the work discussed is not placed in its true perspective. In fact those not actively engaged in this field, and it is such for whom this book appears to

be primarily intended, could not do better than to read Frank's review article (*Advances in Physics* (*Phil. Mag. Suppl.*), 1952, **1**, 91) before starting to read the present work.

Despite this lack of perspective this work does fulfil a need in gathering together, for the first time, an account of much recent experimental work: this is its main virtue. Unfortunately its weaknesses are many. There are far too many obscurities, errors, and dubious statements for these to be mentioned individually. The accounts of work other than the author's are largely précis of the original papers and in fact there is a great deal of quotation, either word for word or almost so, without the use of quotation marks. This is doubly unfortunate in that such accounts are largely reliable, whereas the author's comments or interpolations, throughout the whole of the book, can never be taken on trust, but must always be subjected to critical scrutiny. This may be no bad way of reading any book, but it is hardly to be commended for any but the specialist reader. To give one example, the astonishing statement is made in connection with the distinction between right- and left-handed screw dislocations that "since these dislocations may be turned into one another by turning the crystal, this distinction would again be trivial but....". Apparently the author does not screw nuts on to bolts!

The key to the experimental approach to the subject lies in the optical techniques, but the chapter on modern optical techniques is far from satisfactory. The breathing technique and phase contrast microscopy had been used for the study of crystal surfaces by several workers before the author, despite the statements made. The treatment of phase contrast microscopy is unsatisfactory and inadequate, the reader being presented with a very simplified theory which apparently bears no relation to the experimental results presented. One gathers that phase plateaux on crystals will be revealed by differences in intensity in the image, but the pictures of silicon carbide crystals do not appear to show a single example of this kind. Instead the plateaux are revealed by the steps at their edges, this type of step imaging being a common phenomenon in phase microscopy, although this is not mentioned by the author. (It is relevant to mention at this point that many of the author's pictures of silicon carbide crystals show rather obtrusive markings, presumably caused by the presence of dust on the specimen or in the optical system of the microscope. No mention of this has been seen in the text although this will surely prove confusing to the reader unaccustomed to photomicrography, who has to search for features of low visibility on these pictures.)

One remarkable point is that although it is stated that a complete list of acknowledgments of illustrations is given, this list refers almost entirely to the author's own pictures. Of the large proportion of diagrams either reproduced directly or re-drawn, and also of photographs, no acknowledgment is made to either the original author or publisher, and in fact the reader is left to infer that all but four of the diagrams, and all the photographs, are the work of the author. This is highly discourteous on the part of both author and publisher.

In brief one cannot call this a good book, but when read with critical attention it is a useful book.

L. J. GRIFFIN.

CONTENTS OF SECTION A

PAGE

The late Dr. W. D. Brodie. Lens Spectrometer Study of the Disintegration of ^{233}Pa 397

Prof. S. Devons and Dr. G. Goldring. Emission of Electron–Positron Pairs from Light Nuclei: II—γ-Transitions in ^{8}Be, ^{10}B and ^{16}O 413

Dr. P. R. J. Burch. Cosmic Radiation: Ionization Intensity and Specific Ionization in Air at Sea Level 421

Dr. P. R. J. Burch. Statistical Errors at Background Intensities in Integrating Ionization Chambers 431

Dr. S. E. Hunt, Mr. W. M. Jones, Mr. J. L. W. Churchill and Mr. D. A. Hancock. The Determination of the Resonant Energies for Proton Capture by ^{24}Mg and ^{25}Mg below 550 kev, and Measurement of the Half Lives of ^{25}Al and ^{26}Al 443

Dr. F. M. Kelly, Dr. H. Kuhn and Lady Anne Pery. Hyperfine Structures in the Atomic Spectrum of Calcium 450

Dr. B. L. Moiseiwitsch and Dr. A. L. Stewart. Approximate Molecular Orbitals: II—The $2p\pi_u$ and $3d\pi_g$ States of H_2^+ 457

Research Notes:

Dr. N. H. March and Dr. B. Donovan. Free Electron Diamagnetism and Susceptibilities of the Alkali Metals 464

Mr. G. C. Reid. Energy Levels in ^{10}B and ^{8}Be 466

Prof. E. W. Titterton and Mr. T. A. Brinkley. Cross Sections for the Reaction $^{7}Li(\pi p)^{6}He$ at 17·6 and 14·8 Mev and the First Excited State of ^{6}He 469

Mr. G. R. Satchler. Some Remarks on Radiative Capture and Stripping Reactions 471

Dr. A. G. Gaydon and Mr. A. R. Fairbairn. Pressure Broadening in the Spectrum of NO and its Photodissociation 474

Dr. L. F. H. Bovey and Mr. W. R. S. Garton. The Absorption Spectrum of Thulium 476

Letters to the Editor:

Mr. W. R. Hindmarsh, Dr. H. Kuhn and Mr. S. A. Ramsden. Isotope Shifts in the Atomic Spectra of Tin and Cadmium 478

Dr. S. E. Hunt, Mr. W. M. Jones and Mr. J. L. W. Churchill. Measurement of the End Point Energy of Positrons from ^{25}Al 479

Mr. I. S. Hughes and Dr. P. J. Grant. Polarization of the 10·4 Mev Gamma-Ray in the Reaction $^{27}Al(p\gamma)^{28}Si$ 481

Reviews of Books 482

Contents of Section B 484

Adiabatic Theory of an Electron Gun for Crossed Field Devices

By J. DAIN and I. A. D. LEWIS
Atomic Energy Research Establishment, Harwell, Berkshire

MS. received 21st August 1953 and in amended form 9th March 1954

Abstract. An electron gun is described capable of producing a ribbon-shaped beam of electrons moving under crossed electric and magnetic fields.

The characteristics of the gun are derived by applying the adiabatic principle to motion in constant uniform crossed fields.

§ 1. Introduction

In designing electron guns it is usually required to find the resultant motion of the particles in a region of constant uniform field which may be purely electric, purely magnetic, or some combination of the two. The principle of adiabatic invariance, first discovered in a special case by Lord Rayleigh, and later generalized by Ehrenfest (1914), may be applied to a periodic dynamical system in which the parameters vary slowly with time. The case of a charged particle moving in inhomogeneous fields may be similarly treated provided the spatial changes in the fields appear to the particle to be brought about slowly.* We are thus able to deduce the final motion of an electron beam in an exit region of constant uniform fields without solving the transitory phase in any time- or space-changing fields between the emitter and the exit region; a precise knowledge of the intermediate fields is not required provided they vary slowly. Furthermore, a general solution of this nature may show clearly the dependence of the exit beam characteristics on the various electrical and geometrical parameters. In these circumstances the design may be readily adjusted to suit individual requirements.

In this paper we shall consider the particular case of a gun forming a beam in crossed electric and magnetic fields. The adiabatic invariant of the motion in constant uniform crossed fields is used to obtain details of the beam. The discussion has been limited to the two-dimensional case where the electric field is derived from a potential of the form Ey with E constant, but the analysis may be applied equally well with the field derived from a potential Er. This type of radial field is approximately that found between two coaxial cylinders with a spacing between them small compared with the radius of the inner cylinder. When this restriction is removed an invariant may still be found, but its precise form is not readily obtained.

§ 2. Fundamental Principles

Consider a particle of mass m and charge e moving under the influence of a constant uniform electric field $\mathbf{E}$ and a constant uniform magnetic field $\mathbf{B}$ normal to $\mathbf{E}$. Parallel to the magnetic field the particle is subject to no force and has a constant velocity. Hence, without loss of generality, the problem

* A similar perturbation method is discussed by Alfvén (1950).

will be treated as a two-dimensional one in the plane normal to the magnetic field. In this plane the motion is well known to be that generated by a point on the circumference of a circle radius r rotating with angular velocity ω while its centre moves with a velocity $\mathbf{V}$ where

$$\omega = -(e/m)\mathbf{B}, \qquad \mathbf{V} = (1/B^2)\mathbf{E} \times \mathbf{B}, \qquad \omega^2 r^2 = (\mathbf{v}_0 - \mathbf{V})^2, \quad \ldots\ldots(1)$$

$\mathbf{v}_0$ being the velocity of injection. For electrons, e of course must be reckoned negative.

If now the fields acting on the particle are varied slowly, that is to say in a period long compared with the period of revolution $2\pi/\omega$ and so that neither $\mathbf{E}$ nor $\mathbf{B}$ vary significantly in a distance of the order of r, then the adiabatic principle may be applied to the motion. In the Appendix it is shown that ωr^2 is invariant for such changes in the fields.

The adiabatic change in the fields may be brought about by giving them a time variation. Alternatively it may be caused by the drift velocity $\mathbf{V}$ carrying the particle through a region where the fields have a spatial variation. In this case the spatial change must be sufficiently extended so that fractional changes in the fields, in a distance of the order of r in any direction and in a distance of the order of $2\pi V/\omega$ in the direction of $\mathbf{V}$, shall be small compared with unity. Thus, when V is small and ω large, the changes may be relatively rapid; when V is large and ω small, relatively slow.

Now suppose that the particle starts in one region of constant uniform crossed fields, passes through a region subject to a time or space variation and emerges into a second region of constant uniform crossed fields. Assume that the adiabatic principle may be applied and append suffixes 0 to quantities in the injection region and suffixes 1 in the exit region. Using equations (1), the motion in this latter region is specified completely from the following equations:

$$\left.\begin{aligned} \mathbf{V}_0 &= (1/B_0^2)\mathbf{E}_0 \times \mathbf{B}_0, & \mathbf{V}_1 &= (1/B_1^2)\mathbf{E}_1 \times \mathbf{B}_1, \\ \omega_0 &= -(e/m)\mathbf{B}_0, & \omega_1 &= -(e/m)\mathbf{B}_1, \\ \omega_0^2 r_0^2 &= (\mathbf{v}_0 - \mathbf{V}_0)^2, & \omega_1 r_1^2 &= \omega_0 r_0^2. \end{aligned}\right\} \quad \ldots\ldots(2)$$

The nature of the fields in the transition region is unspecified beyond the restriction that the change should appear adiabatic to the particle. If a spatial variation of the fields is required it can be produced by suitably shaped electrodes and magnet pole pieces.

Neither $\mathbf{V}$ nor ω necessarily remains constant during the motion, but the conditions imposed by the term 'adiabatic' may be found in orders of magnitude by a rough averaging process. This is adequate for practical purposes. Thus, for time-varying fields, the transition must take place in a period long compared with $4\pi/(\omega_0 + \omega_1)$. For fields varying smoothly in space over a distance a then

$$a \gg \tfrac{1}{2}(2\pi V_0/\omega_0 + 2\pi V_1/\omega_1). \quad \ldots\ldots(3)$$

Since v_0 is generally not an order of magnitude greater than V_0 this condition is sufficient to satisfy both restrictions placed on the fractional changes in the fields.

The criterion (3) may be expressed differently as follows. Let r' be the generating circle radius for zero velocity injection into fields $\mathbf{E}$ and $\mathbf{B}$, then, from relations (1),

$$r' = V/\omega \quad \ldots\ldots(4)$$

and the condition becomes

$$a \gg \pi(r_0' + r_1'). \quad \ldots\ldots(5)$$

§3. An Adiabatic Electron Gun

Suppose electrons are emitted with zero velocity from a cathode situated in constant uniform crossed fields $\mathbf{E}_0$ and $\mathbf{B}_0$. Let the electrode and magnet pole assembly be such that the electrons drift through a transition field region into a second region of constant uniform crossed fields $\mathbf{E}_1$, $\mathbf{B}_1$, where the beam is to be formed. In the transition either or both of the electric and magnetic fields may be changed. The characteristics of the output beam are readily deduced, neglecting space-charge forces.

Take axes such that Ox lies along the cathode in the direction of the drift velocity. Then, using the previous notation,

$$\left.\begin{aligned} r_0 = r_0' = V_0/\omega_0, \qquad & \omega_1 r_1^2 = \omega_0 r_0^2, \\ \omega_0 = |e/m|B_0, \qquad & V_1 = E_1/B_1, \\ \omega_1 = |e/m|B_1. & \end{aligned}\right\} \quad \ldots\ldots(6)$$

The electron velocities lie within the limits $U_{\max}$ and $U_{\min}$ where

$$U_{\max} = V_1 + \omega_1 r_1, \qquad U_{\min} = V_1 - \omega_1 r_1, \quad \ldots\ldots(7)$$

and the beam has a breadth $2r_1$ normal to the drift velocity.

Elimination between equations (7) and (6) yields

$$\left.\begin{aligned} U_{\max} = V_1(1+f), \qquad & U_{\min} = V_1(1-f), \\ 2r_1 = 2(V_1/\omega_1)f, & \\ \text{with} \qquad f = (E_0/E_1)(B_1/B_0)^{3/2}. & \end{aligned}\right\} \quad \ldots\ldots(8)$$

When the fields are conservative it is seen that the electrons oscillate between a maximum and minimum displacement from the cathode, $y_{\max}$ and $y_{\min}$, given by the relations

$$|e|E_1 y_{\max} = \tfrac{1}{2}mU_{\max}^2, \qquad |e|E_1 y_{\min} = \tfrac{1}{2}mU_{\min}^2. \quad \ldots\ldots(9)$$

It is convenient here to introduce the concept of critical magnetic field in the main beam region. This is the magnetic field required just to turn electrons, emitted in a uniform field, away from the anode. This field, B_{cr}, is related to the anode–cathode separation d and the electric field E_1 by the equation

$$E_1 = \tfrac{1}{2}|e/m|B_{cr}^2 d. \quad \ldots\ldots(10)$$

Combining equations (8), (9) and (10) we obtain

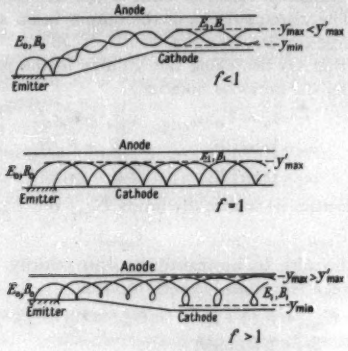
$$\left.\begin{aligned} y_{\max}/d &= \tfrac{1}{4}(B_{cr}/B_1)^2(1+f)^2, \\ y_{\min}/d &= \tfrac{1}{4}(B_{cr}/B_1)^2(1-f)^2, \\ 2r_1/d &= (B_{cr}/B_1)^2 f. \end{aligned}\right\} \quad \ldots\ldots(11)$$

Setting $f=1$ in the above expressions gives the well-known results for the case when the fields are everywhere uniform (i.e. parallel electrodes in uniform magnetic field). Denoting the various parameters in this case by dashed symbols the following results are obtained:

$$y_{\max}/y'_{\max} = \tfrac{1}{4}(1+f)^2, \qquad r_1/r_1' = f. \quad \ldots\ldots(12)$$

Examining (8), (11) and (12) it will be seen that if f is less than unity the beam is narrower than for the case of parallel electrodes in a uniform field; it does not extend down to the cathode ($y_{\min}>0$) and the orbits contain no closed loops ($U_{\min}>0$). When f is greater than unity the beam is broad, does not extend down to the cathode ($y_{\min}>0$) but extends well towards the anode ($y_{\max}>y'_{\max}$). The orbits in this case contain closed loops since $U_{\min}$ is negative. These features are illustrated in the figure.

In practice narrow beams of electrons are required so that f must be made small. Generally speaking it is more convenient to arrange the magnetic field to be uniform throughout the gun; in this case $r_1=r_0$, $f=E_0/E_1$ and a narrow beam gun requires a cathode situated in a weak electric field. There does not appear, however, to be any immediate limitation on the size of cathode used.



Orbits in the adiabatic electron gun.

$$f=\frac{E_0}{E_1}\left(\frac{B_1}{B_0}\right)^{3/2}.$$

§4. Linear Beam Guns

From the invariance of ωr^2 throughout the adiabatic region of the gun proposed here it follows that if r_0 can be made zero then r will always be zero and the resultant beam will be a linear one.

4.1. *Finite Velocity Injection*

From equation (2) we have

$$r_0^2=(1/\omega_0^2)(\mathbf{v}_0-\mathbf{V}_0)^2; \qquad \ldots\ldots(13)$$

thus $r_0=0$ if the electrons are injected with a finite velocity such that $\mathbf{v}_0=\mathbf{V}_0$. The resultant beam will have zero breadth and linear trajectories along the line

$$y=\tfrac{1}{4}(B_{cr}/B_1)^2. \qquad \ldots\ldots(14)$$

4.2. *Space Charge Limited Injection*

From equation (13) it follows that if the injection velocity $\mathbf{v}_0$ and the electric field at the cathode $\mathbf{E}_0$, and hence $\mathbf{V}_0$, are both zero, then r_0 is also zero.

For a cathode that is space charge limited these conditions apply. Hence if the motion of the electrons can be controlled so that the space charge along the beam, and consequently the electric field, varies adiabatically, then the resultant beam should be linear. Furthermore, the electron flow should conform to smooth streamlines and the adiabatic condition on the fields will be required to hold only in the direction of the streamlines. There is no restriction on the variation of the fields transverse to the streamlines, and in fact the resultant beam

can be produced, for example, in an electric field that is non-uniform in a direction normal to the electron flow.

It may be possible to set up laminar flow (Brillouin 1943, 1951) in this manner from a finite emitting cathode provided electronic interaction, such as the 'slipping stream effect' (Bunemann 1950, MacFarlane and Hay 1950), plays a negligible part. It must be mentioned that such a gun has been envisaged previously (Guénard and Huber 1952) but the angle of attack has been different. The invariance principle used here is felt to give a better physical insight into the mode of operation of this type of gun.

§ 5. An Example of Design

This example is quoted solely as an illustration and is not intended as an optimum design. The formulae given in previous sections refer to the m.k.s. system of units. In the following calculations the quantities have been expressed in units more familiar in the laboratory.

Suppose a gun is required to produce a beam of electrons travelling with a velocity one-fifth that of light in a region between two parallel plates, separated by a distance of 1 cm, with an applied potential difference of 60 kv. Let the gun consist of a cathode spaced 3 cm from the anode tapering slowly down to 1 cm separation in the exit space (see figure, $f<1$). The magnetic field is the same everywhere and has the value 1000 gauss to give the required drift velocity. With this geometry, and neglecting space charge forces, we have $f=\frac{1}{3}$, $r_0=r_0'=0{\cdot}113$ cm, $r_1'=r_0'/f=0{\cdot}34$ cm, $(B_{cr}/B_1)^2=2r_1'/d=0{\cdot}68$. The beam thus has the characteristics $y_{min}=0{\cdot}076$ cm, $y_{max}=0{\cdot}302$ cm, $2r_1=0{\cdot}226$ cm, whereas a beam formed between plane parallel electrodes 1 cm apart has $y_{min}=0$, $y_{max}=2r_1'=0{\cdot}68$ cm. The advantage of the 'adiabatic' gun is thus demonstrated.

Applying the criteria for the length of the tapered section as discussed in § 2, since $\pi(r_0'+r_1')=1{\cdot}4$ cm, the length of taper for this particular example requires to be of the order of 10 cm.

§ 6. Conclusion

An electron gun has been outlined the performance of which has been readily predicted (neglecting space charge effects) by applying the adiabatic principle to motion in time-invariant crossed fields.

The orbits have been calculated exactly by the authors and by Poritzsky and Jerrard (1952) in the particular case of a structure consisting of a plane cathode and an anode shaped in the form of a rectangular hyperbola (or vice versa). The results obtained agree with those based on the adiabatic principle subject to the condition of adequate slowness of variation being fulfilled. Experiment has shown, however, that this condition is not very stringent in practice.

Acknowledgments

This paper is published by permission of the Director, Atomic Energy Research Establishment.

References

Alfvén, H., 1950, *Cosmical Electrodynamics* (Oxford : Clarendon Press), pp. 14–34.

Brillouin, L., 1943, *Phys. Rev.*, **63**, 127 ; 1951, *Advances in Electronics*, Vol. III (New York: Academic Press).

Bunemann, O., 1950, *Nature, Lond.*, **165**, 474.

EHRENFEST, P., 1914, see BORN, M., *Atomic Physics*, 1937, 2nd Edn (London: Blackie and Son), p. 101 *et seq.*
GUÉNARD, P., and HUBER, H., 1952, *Ann. Radioélect.*, **7**, 252.
MACFARLANE, G. G., and HAY, H. G., 1950, *Proc. Phys. Soc.* B, **63**, 409.
PORITZSKY, H., and JERRARD, R. P., 1952, *J. Appl. Phys.*, **23**, 928.

APPENDIX

The Invariant of Motion in Crossed Fields

The non-relativistic Hamiltonian is given by

$$H=(1/2m)(\mathbf{p}-e\mathbf{A})^2+e\phi=\mathscr{E} \text{ the total energy}$$

where ϕ is the electric potential, $\mathbf{A}$ the magnetic vector potential and $\mathbf{p}$ the generalized momentum. If the fields are uniform such that, in cartesian coordinates, $\mathbf{E}=(0,\ -E,\ 0)$, $\mathbf{B}=(0,\ 0,\ -B)$, then the corresponding potential functions are $\phi=Ey$, $\mathbf{A}=(By,\ 0,\ 0)$, and $\mathbf{p}$ is obtained from $\mathbf{p}=m\mathbf{v}+e\mathbf{A}$.

Motion in the x, y plane will be irrotational and the action function W may be defined by grad $W=\mathbf{p}$.

Now the Hamiltonian is independent of x so that p_x is constant. Thus, take $\partial W/\partial x=p_0$, a constant, and the Hamilton–Jacobi equation becomes

$$(p_0-eBy)^2+\left(\frac{\partial W}{\partial y}\right)^2+2meEy=2m\mathscr{E}. \qquad \text{.....(A 1)}$$

After re-arrangement this becomes

$$\frac{\partial W}{\partial y}=eB\left\{\frac{2m\mathscr{E}}{e^2B^2}+\frac{m^2E^2-2p_0mEB}{e^2B^4}-\left[\frac{p_0B-mE}{eB^2}-y\right]^2\right\}^{1/2}.$$

Now $\partial W/\partial y$ must be real, so y oscillates between two limits $y_{\max}$ and $y_{\min}$, where

$$\tfrac{1}{4}(y_{\max}-y_{\min})^2=(2m\mathscr{E}/e^2B^2)+[(m^2E^2-2p_0mEB)/e^2B^4].$$

The action integral J, defined by

$$J=\oint\frac{\partial W}{\partial y}\,dy$$

taken over a period in y, has the value

$$J=\tfrac{1}{4}\pi eB(y_{\max}-y_{\min})^2.$$

This integration is obtained by observing that the plot in phase space of $(1/eB)\partial W/\partial y$ against y is a circle of radius $\frac{1}{2}(y_{\max}-y_{\min})$.

By the adiabatic principle J is invariant for slowly changing fields hence $\omega(y_{\max}-y_{\min})^2$ is also invariant.

In constant uniform crossed fields the motion of the particle is that of a point on the circumference of a circle radius r rotating at an angular velocity $\boldsymbol{\omega}$ whilst its centre moves with a velocity $\mathbf{V}$, where $\mathbf{V}=(1/B^2)\mathbf{E}\times\mathbf{B}$, $\boldsymbol{\omega}=-(e/m)\mathbf{B}$. From this it follows that $y_{\max}-y_{\min}=2r$. Hence ωr^2 is invariant for adiabatic changes in the fields $\mathbf{E}$ and $\mathbf{B}$.

It now remains to establish its value. Suppose that particle is injected when $t=0$ at the point $(x_0,\ y_0)$ with a velocity $\mathbf{v}_0$ in fields $\mathbf{E}_0$ and $\mathbf{B}_0$ where

$$\mathbf{v}_0=(dx_0/dt,\ dy_0/dt,\ 0),$$
$$\mathbf{E}_0=(0,\ -E_0,\ 0), \qquad \mathbf{B}_0=(0,\ 0,\ -B_0),$$

then, inserting these values gives

$$\mathscr{E} = \tfrac{1}{2}m\left\{\left(\frac{dx_0}{dt}\right)^2 + \left(\frac{dy_0}{dt}\right)^2\right\} + eE_0y_0,$$

$$p_0 = m\frac{dx_0}{dt} + eB_0y_0.$$

Further substitution yields

$$\begin{aligned} r_0^2 &= (2m\mathscr{E}/e^2B_0^2) + (m^2E_0^2 - 2p_0mE_0B_0)/e^2B_0^4 \\ &= \left(\frac{m}{eB_0}\right)^2\left[\left(\frac{dx_0}{dt}\right)^2 + \left(\frac{dy_0}{dt}\right)^2 + \frac{2e}{m}E_0y_0 + \left(\frac{E_0}{B_0}\right)^2 - 2\frac{E_0}{B_0}\frac{dx_0}{dt} - 2\frac{e}{m}E_0y_0\right] \\ &= (m/eB_0)^2[(dx_0/dt - E_0/B_0)^2 + (dy_0/dt)^2]. \end{aligned}$$

Thus the final result is

$$\omega r^2 = \omega_0 r_0^2 = (1/\omega_0)(\mathbf{v}_0 - \mathbf{V}_0)^2.$$

Effect of Grain Size on the Crystal Structure of Cobalt

By E. A. OWEN and D. MADOC JONES
University College of North Wales, Bangor

MS. received 29th January 1954

Abstract. The paper gives an account of measurements made on the lattice parameter of cobalt when the material contained crystal grains of different sizes as found in fine dust, filings and annealed solid rod. The stable structure of cobalt was found to depend upon the grain size: between room temperature and about 450°C when the grain size is very small, as in cobalt sponge, the stable structure is face-centred cubic; when the grain size is larger, as in a solid rod, the stable structure is close-packed hexagonal; when there is a range of grain size, as in fine filings a mixture of these two structures is observed. At temperatures above 450°C cobalt sponge is still face-centred cubic, and the filings and rod also now show face-centred cubic structure with the same lattice parameter (3·5370 kx at 18°C). The lattice constants of the hexagonal close-packed structure at 18°C are: $a = 2{\cdot}5003$ kx, $c/a = 1{\cdot}632_2$.

Different results were obtained according as the material was maintained and examined at elevated temperature or was quenched from elevated temperature and examined at room temperature. All specimens of cobalt, whatever the grain size, annealed at about 1000°C and quenched showed a mixture of the two structures although with cobalt sponge the amount of the hexagonal structure was very small compared with the amount of the cubic structure.

Combining these results with recent results by Newkirk and Geisler (1953) it is possible to present a fairly complete picture of the effects observed, between room temperature and 1220°C, with material examined at elevated temperature and with quenched material examined at room temperature. The main results fit the theory of the imperfect hexagonal structure put forward by Edwards and Lipson (1943).

A cubic face-centred structure with lattice parameter (3·5540 kx) greater than that of the cubic lattice ordinarily found (3·5370 kx) was observed with filings quenched from temperatures between about 600°C and 840°C. It is suggested that the material giving this lattice parameter is in a metastable state.

The thermal expansion of cobalt sponge is represented by a smooth curve the equation of which, for temperatures up to 650°C, is the quadratic $a_t = \alpha_0(1 + \alpha t + \beta t^2)$, where $a_0 = 3{\cdot}5362$ kx; $\alpha = 12{\cdot}297 \times 10^{-6}$; $\beta = 2{\cdot}042 \times 10^{-9}$. It shows no transformation such as is found at about 450°C with material in solid form. No other high-temperature transformation was observed.

§1. Introduction

Much has been written on the crystal structure of cobalt and many measurements have been made on the lattice parameters of the two allotropic forms in which it exists, and of the temperature at which the transformation from one structure to the other occurs, but there appears to be

much variation in these values determined by different workers, as will be observed from the lattice parameter values quoted in table 1. There are also variations in the statements made concerning the existence and the actual values of the transformation temperatures; both high-temperature and low-temperature transformations have been recorded. Latterly attention has been directed to the effect of grain size on these temperatures and also on the structures that are stable at different temperatures; more information is, however, desirable on these points.

The work described in this paper was undertaken with a view to supplying further information about the behaviour of cobalt when examined by x-ray diffraction at different temperatures, and about its behaviour when examined in different forms, varying from very fine dust to solid material in wire form, thereby to gain further knowledge of the effect of grain size on the crystal structure.

It is interesting to note in table 1 that the value of the cubic parameter of cobalt found at room temperature by Sekito (1927) and by Marick (1936) and that quoted by Wyckoff (1931) is about $3{\cdot}557 \pm 0{\cdot}002$ kx, but that later values found by Taylor (1950) and by Ellis and Greiner (1948) are definitely smaller, agreeing more closely with the value quoted by Neuburger (1936). Similar variations, perhaps not so marked, are also found in the values of the hexagonal lattice

Table 1. Values of Lattice Parameters of Cobalt at 18°C

Author	Lattice parameter (kx units) Cubic lattice	Hexagonal lattice a	c/a
Sekito (1927)	$3{\cdot}558 \pm 0{\cdot}002$	2·498	1·622
Neuburger (1936)	3·54(5)	$2{\cdot}50_7 \pm 0{\cdot}005$	1·624
Marick (1936)	$3{\cdot}559 \pm 0{\cdot}002$	$2{\cdot}519 \pm 0{\cdot}002$	1·624
Wyckoff (1931)	$3{\cdot}554 \pm 0{\cdot}002$	$2{\cdot}514 \pm 0{\cdot}002$	1·633
Taylor (1950)	3·5369	2·5013	$1{\cdot}623_7$
Ellis and Greiner (1948)	3·5368	2·5020	$1{\cdot}623_1$

parameters. It should be mentioned that the values given by Taylor and by Ellis and Greiner were not known to us until after the completion of the work on cobalt, so that more work was done than was really necessary because our results differed so markedly from those known to us at the time. It was gratifying to us to find that our results corroborated the later values.

§ 2. Apparatus

The investigation was carried out with the aid of two cameras, one being a new high-temperature focusing camera with which a series of x-ray photographs of reflections at high Bragg angles could be obtained without disturbing the specimen, the other being the ordinary Debye–Scherrer type for use at room temperature. Access was had also to another focusing camera suitable for room-temperature examinations and provided with a mechanism to oscillate the powder specimen along the circumference of the camera, thus making it possible to obtain more satisfactory reflections that could be accurately measured. The new high-temperature focusing camera will be described in detail elsewhere. Some trouble was experienced in measuring the temperature of the specimen to the accuracy desired; the final arrangement adopted yielded entirely satisfactorily results, the measured temperatures being well within $\pm 1{\cdot}0\%$ of the true temperature, as calculated from well-established thermal curves of pure standard substances.

§3. Material

For preliminary work with the new high-temperature camera cobalt of purity not less than 99·9% was used. It was employed in the form of fine filings taken from a lump of the material. Thermal expansion curves were obtained with this material which later were found to differ little from those found with purer material. The main object of using this material at this stage was to gain experience in the technique of temperature measurement with the new arrangement and in the general manipulation of the apparatus. Afterwards, for the main investigation, the very purest cobalt obtainable from Messrs. Johnson Matthey and Co. was procured. It was examined in three forms: fine dust as supplied (cobalt sponge); wire, 1 mm diameter, made from the sponge by Messrs. Johnson Matthey and estimated by them to contain 0·004% of iron; and filings taken from the wire with a fine file. To ensure that no iron impurity is introduced, as might be the case if an ordinary file were used, a metal-bonded diamond hand-hone was employed to produce the filings. All the cobalt specimens when they were being annealed were sealed in evacuated silica containers.

§4. Experiment

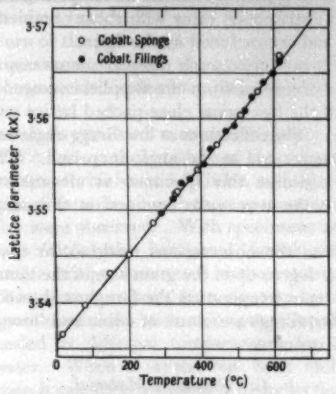
It will be convenient to consider the experimental work in two sections, namely, that carried out when the material was maintained at elevated temperatures and that carried out with quenched material.

(i) *Material maintained at Elevated Temperature*

Cobalt sponge. This material was given an initial heat treatment by annealing for 12 hours at a temperature between 520°c and 550°c and quenched in iced water. An x-ray photograph taken with manganese Kα radiation at room temperature on the rotating-fibre camera showed no trace of the presence of a hexagonal structure; the reflections were well defined and the x-ray pattern represented a face-centred cubic structure with a lattice parameter of $3{\cdot}5370 \pm 0{\cdot}0002$ kx at 18°c. Another sample of the same material after a similar treatment yielded again a pure face-centred cubic structure without admixture with a hexagonal structure, the value of the lattice parameter being 3·5369 kx. These values did not agree with the early values shown in table 1 and known to us at the time the measurements were made.

A series of x-ray photographs was now taken with one of the above specimens when it was maintained at various temperatures between room temperature and 620°c in the high-temperature focusing camera. The material yielded good reflections over the whole temperature range; the α_1 and α_2 components of the (311) reflection were sharp and well resolved throughout and could be accurately measured. At the lower temperatures, after the material had been maintained at the highest temperatures, a very faint doublet at a slightly higher Bragg angle than the strong (311) reflection was just visible; these faint lines were identified as the $(11\bar{2}2)$ reflections from the hexagonal lattice. They were too faint to be measured and their presence did not interfere with the accurate measurement of the reflections from the cubic lattice. The intensity of the (311) reflections did not change visibly as the temperature was altered, and the almost complete transformation from the cubic to a hexagonal structure mentioned by Edwards and Lipson (1943) was not observed, the only indication of a hexagonal structure being the very faint reflections already mentioned. It is significant that with this

fine-grain material only the face-centred cubic lattice exists whether the material is maintained at temperatures up to 600°C or quenched from these temperatures. Accurate measurements of the reflections were possible, and these yielded the



thermal expansion curve for the cubic or β phase of cobalt shown in the figure, the equation to which, worked out in the usual manner, is found to be $a_t = a_0(1 + \alpha t + \beta t^2)$, where $a_0 = 3{\cdot}5362\,\text{kx}$; $\alpha = 12{\cdot}297 \times 10^{-6}$; $\beta = 2{\cdot}042 \times 10^{-9}$. The true coefficients of expansion are given in table 2. These figures refer to the expansion of the pure cubic lattice.

Table 2. Coefficients of Thermal Expansion of the Cubic Lattice of Pure Cobalt Sponge

t (°c)	$\frac{1}{a_t}\frac{da}{dt} \times 10^6$	$\frac{1}{a_0}\frac{da}{dt} \times 10^6$	$\frac{a_t - a_0}{a_0 t} \times 10^6$
0	$12{\cdot}3_0$	$12{\cdot}3_0$	—
50	$12{\cdot}4_9$	$12{\cdot}5_0$	—
100	$12{\cdot}6_9$	$12{\cdot}7_1$	$12{\cdot}4_4$
150	$12{\cdot}8_9$	$12{\cdot}9_1$	—
200	$13{\cdot}0_8$	$13{\cdot}1_1$	$12{\cdot}6_5$
250	$13{\cdot}2_8$	$13{\cdot}3_2$	—
300	$13{\cdot}4_7$	$13{\cdot}5_2$	$12{\cdot}9_3$
350	$13{\cdot}6_7$	$13{\cdot}7_2$	—
400	$13{\cdot}8_6$	$13{\cdot}9_3$	$13{\cdot}1_5$
450	$14{\cdot}0_5$	$14{\cdot}1_3$	—
500	$14{\cdot}2_4$	$14{\cdot}3_4$	$13{\cdot}3_8$
550	$14{\cdot}4_4$	$14{\cdot}5_4$	—
600	$14{\cdot}6_3$	$14{\cdot}7_5$	$13{\cdot}5_5$

Cobalt filings. The filings were taken from material of purity not less than 99·9%; these were given a preliminary heat treatment of 2 hours at 500°C followed by quenching in iced water. The x-ray pattern showed traces of the hexagonal lattice mixed with the cubic lattice, and thus differed from the pattern obtained with the cobalt sponge. X-ray photographs taken with the material when maintained at temperatures in the range from 625°C to 300°C yielded reflections

from the cubic lattice which could be measured with fair accuracy. But below about 300°c the high-angle reflections became less well defined, and at the lower temperatures degenerated into a diffuse band which could not be measured. The values of the cubic lattice parameter at temperatures between 300°c and 625°c agreed within experimental error with those obtained with the fine cobalt dust over this range and are included in the figure.

Cobalt wire. An x-ray photograph taken at room temperature with the wire in the condition it was received from the suppliers showed a pattern which was predominantly that of the hexagonal close-packed lattice with faint traces of the cubic (200) reflection. The reflections at low Bragg angle were strong and sharp, but the definition deteriorated as the angle increased. There was no suitable camera available to examine this specimen at elevated temperatures so the observations made on the wire were confined at this stage to those at room temperature.

It was evident from the observations made so far that the x-ray patterns depended to a marked degree upon the grain size in the material when submitted to examination. At room temperature the fine dust showed only a face-centred cubic structure, the fine filings a mixture of cubic and hexagonal structures, and the solid a hexagonal structure..

(ii) *Quenched Material*

The investigation was continued with cobalt in various forms quenched after being annealed for different periods at elevated temperatures; the filings in this part of the investigation were taken from the wire. The purity of the material was therefore the same throughout except that the wire, and therefore the filings, contained 0·004% of iron. It was considered that this amount of impurity was insufficient to prevent a fair comparison being made between the results obtained with the sponge and with the other cobalt specimens. A number of specimens of the three forms of cobalt were accordingly given various heat treatments and their x-ray patterns obtained with a standardized Debye–Scherrer camera employing throughout manganese Kα radiation as before. Some lattice parameters were also calculated from the measurements of reflections recorded in a focusing camera furnished with an oscillating mechanism.

Cobalt sponge. This material, after being maintained at 966°c for one hour and then quenched, yielded an x-ray pattern showing a cubic lattice with a very small admixture of a hexagonal lattice. The same specimen held for 10 hours at 500°c and slowly cooled to room temperature at the rate of about 40°c per hour yielded the same pattern. The hexagonal reflections were strong enough for measurement; the axial ratio c/a was found to be 1·633 and the a parameter at 18°c $2{\cdot}501 \pm 0{\cdot}003$ kx. The cubic lattice parameter was the same as before, namely 3·5370 kx at 18°c. Slight errors in the measurements of the films may have been caused by the close proximity of the cubic to some of the hexagonal reflections used to calculate the parameters. The lines involved were sharply defined, but if the reflections from the two lattice structures did not exactly coincide, the measurements of the hexagonal reflections would be influenced by the more intense reflections from the cubic lattice.

Another specimen of cobalt sponge annealed at 866°c for 2 hours and quenched gave a pattern showing predominance of cubic with a small amount of hexagonal structure. When, however, a specimen of the material was annealed at 356°c

for 7 days, and again at 327°c for 5 days, and afterwards quenched, the pattern was almost entirely cubic with only traces of the hexagonal reflections showing. The same condition resulted after 20 hours at 519°c; a photograph taken with this specimen on the focusing camera showed a very well defined (311) reflection only. Similarly another specimen held at 600°c for 21 hours and quenched gave the same pattern, the values of the parameter of the cubic lattice being the same as that found in the previous measurements, viz. 3·5370 kx at 18°c. A sample of the unannealed cobalt sponge yielded reflections that were diffuse and could not be measured.

Summarizing the results obtained with cobalt sponge, samples annealed at temperatures up to 600°c and quenched showed the presence of a face-centred cubic lattice the parameter of which was 3·5370 kx, and this value was found to be independent of the time of annealing; with these specimens only very slight traces of the hexagonal lattice were observed. With specimens annealed at the higher temperatures (866°c to 966°c) the hexagonal lattice was clearly recorded, although the cubic lattice still predominated in the pattern.

Cobalt filings. Filings were prepared from the wire by means of the diamond hand-hone already mentioned. As with the sponge, specimens of the filings, unsieved, were annealed at different temperatures for varying periods and quenched in iced water. When a crystal has been sufficiently annealed to render it free from strain it usually gives sharp x-ray reflections. All the annealed samples of cobalt sponge mentioned above yielded well-defined sharp x-ray reflections from planes in the cubic lattice, the reflection from the (311) planes being well resolved into its α_1 and α_2 components. This, however, was not the case with any of the specimens of filings that were examined. The α_1 and α_2 components of the (311) reflections could now only just be distinguished on some of the films, and in no instance were they as sharp as those found throughout the work on cobalt sponge. It would appear that the crystal lattice formed with this grade of powder never attains the same degree of equilibrium at these temperatures as that obtained with the much finer grains of the sponge.

The results obtained at temperatures below about 450°c showed that the filings contained both cubic and hexagonal structures. Also, when annealed at 966°c for 1 hour and quenched, a mixture of structures was observed, the lattice parameters of both structures being the same as those found with the sponge. No alteration in the appearance of the diffraction pattern occurred when the material was further annealed at 500°c for 10 hours and slowly cooled to room temperature, and still further annealed at about 200°c for 5 days and quenched.

When a sample of filings was annealed at a temperature of 595°c a new effect was observed. The specimen was maintained at this temperature for 2, 5, 24 and 60 hours respectively, and after each interval it was quenched in iced water and its x-ray pattern photographed. The first two photographs showed the presence of a cubic lattice of the same size as the one already observed, with the (311) reflection doublet nearly resolved, and a trace of the hexagonal lattice. The last two photographs showed a deterioration in the definition of the (311) reflections, and at the same time the cubic lattice was progressively increasing in size. The values of the lattice parameter in the four photographs were 3·5367, 3·5366, 3·5375, 3·5400 kx. When the sample was further annealed for 10 hours at 500°c and slowly cooled to room temperature no further change occurred.

Another specimen annealed for 14 hours at 836°c and quenched emphasized the change observed with the previous samples held at 595°c and quenched. The x-ray pattern now showed the presence of a cubic structure entirely free from admixture with hexagonal structure, but having a larger lattice parameter than the cubic lattice previously observed. Annealing the specimen for a further period of 14 hours at 830°c and quenching caused no change in the appearance of the pattern. The lattice parameter was found to be 3·5540 kx at 18°c, a value which is almost identical with the values quoted by Sekito, Marick, and Wyckoff. This specimen was further annealed for periods of 24 and 36 hours respectively and quenched; afterwards it was annealed for 10 hours and slowly cooled to room temperature and finally annealed for 5 days at a temperature ranging between 168°c and 268°c and then quenched, but the pattern remained the same after each of these treatments. The large cubic lattice persisted without a trace of the hexagonal. This specimen had undergone heat treatment which, according to Edwards and Lipson (1943), should have rendered the cobalt 80% hexagonal. This effect was not observed; in fact the x-ray pattern emphasized the absence of the hexagonal structure.

In all the photographs showing the larger cubic lattice the reflections at high Bragg angles were diffuse, the (311) doublet remaining completely unresolved; but although accurate measurements were thereby rendered impossible, sufficient accuracy was possible to show that the lattice had considerably increased in size.

A new sample of filings was annealed at 492°c for 24 hours and quenched. The x-ray pattern showed the presence of the small cubic lattice and a trace of the hexagonal. Further annealing at 496°c for 2 hours and quenching, and 10 hours at 500°c followed by slow cooling to room temperature, did not appreciably alter the pattern. Another specimen annealed at 616°c for 2½ hours and quenched showed an x-ray pattern which was mainly cubic with a lattice parameter of intermediate value between the largest and smallest; similarly another specimen annealed at 650°c for 18 hours and quenched showed an intermediate value of the lattice parameter. These results are in agreement with the previous findings recorded above of the change in parameter of the cubic lattice when maintained at 595°c for long periods of time. When, however, a specimen was maintained at 966°c for 1 hour and quenched, the pattern contained sharp reflections from the small cubic lattice together with reflections from the hexagonal lattice. Further heat treatment at 500°c for 10 hours followed by slow cooling to room temperature, and again 5 days annealing at a temperature varying between 168°c and 268°c, produced no change.

It is noteworthy that the filings when annealed at 966°c sintered; the amount of sintering was reduced, but not entirely removed, by mixing with the filings a small quantity of alumina powder of suitable grade.

Cobalt wire. A Debye–Scherrer photograph taken with a sample of the cobalt wire as received from the suppliers showed reflections which were mainly from the hexagonal lattice, together with a faint trace of the cubic (200) reflection. The reflections at low Bragg angles were strong and well defined; the definition deteriorated as the angle increased. After annealing for 12 hours at 418°c and quenching, the lines in the x-ray pattern showed better definition although the high-angle reflections still remained totally unresolved. No further change was noticed after annealing for 7 days at 418°c, but this photograph showed only very faint traces of the presence of a cubic structure. A photograph of the pure

hexagonal state with the high-angle reflections well resolved was not obtained, the annealing temperature of 418°C being probably too low to remove the lattice strain in a reasonable period of time. These observations confirm those of previous workers (Edwards and Lipson 1943, Hull 1921, Hendricks, Jefferson and Shultz 1930).

The final values of the hexagonal lattice parameters were worked out by the Cohen (1935) analytical extrapolation method from measurements made on the last film described above. Seven different reflections were used; the final values of the parameters at 18°C resulting from the calculations were: $a = 2{\cdot}5003_2 \pm 0{\cdot}0003$ kx; $c = 4{\cdot}0809_6 \pm 0{\cdot}0003$ kx; $c/a = 1{\cdot}632_2$. These figures are considered more accurate than those quoted earlier in the paper and are to be preferred because the hexagonal reflections were free from interference by the superposition of cubic reflections.

When a specimen of the wire was annealed at 514°C for 36 hours and quenched, the structure present was predominantly the smaller cubic lattice with visible traces of the hexagonal $(10\bar{1}1)$ reflection. Very little change was observed when the wire was annealed for a further period of 5 hours at 510°C and slowly cooled to room temperature at the rate of about 40°C per hour. The reflections from the small cubic lattice were strong and sharp; the hexagonal $(10\bar{1}1)$ reflection was very faintly visible but quite sharp.

When the annealing temperature was raised to 960°C and the specimen maintained at this temperature for 5 hours and quenched, the X-ray photograph contained clear reflections from both the cubic and the hexagonal lattices. The (311) cubic and the $(11\bar{2}2)$ hexagonal reflections could be just distinguished, the latter being at a slightly higher Bragg angle than the former. When the structures appeared together the appearance of the X-ray reflections was as described in table 3.

Table 3. Description of Film when Cubic and Hexagonal Reflections appear together

Reflection	Description of appearance of reflection
$10\bar{1}0$	Fairly strong and sharp.
0002	Strong and sharp; coincides with (111) cubic.
$10\bar{1}1$	Strong and sharp.
$10\bar{1}2$	Very very faint but sharp.
$11\bar{2}0$	Strong and sharp; coincides with (220) cubic.
$10\bar{1}3$	Faint, broad and diffuse.
$20\bar{2}0(1)$, $20\bar{2}0(2)$	Rather fainter than $(10\bar{1}3)$ and totally unresolved.
$11\bar{2}2(1)$, $11\bar{2}2(2)$	Sharp and resolved; coincides with (311) cubic.

This table confirms the observations of Edwards and Lipson (1942) that the hexagonal reflections which coincide with the cubic reflections are sharply defined, whereas those that do not coincide, especially at the higher angles, are broadened, and the phenomenon has been satisfactorily explained by Edwards and Lipson (1942) and by Wilson (1942) on the assumption that the probability that the alternate layers of atoms in the structure are similar, is less than unity. Hendricks *et al.* (1930) found "the reflection due to the planes $(10\bar{1}1)$, which should be the strongest reflection from the hexagonal lattice, was often faint, and markedly so in those specimens reduced below 400°C but still evident in the metal

reduced at 1020°C ". The present results agree with the first part of this statement but differ from the second part, as will be seen from table 3, and all the films on which both the hexagonal and the cubic lattices were present had the characteristics detailed in the table. On looking through the literature on cobalt one encounters other disagreements between the results of different observers, and these arise probably from the fact that they do not all refer to material in the same state as regards grain size. In addition to differences concerning the crystal structures that are stable at different temperatures, there are also variations in the estimate of the temperature at which the transformation from the α to the β phase occurs.

§5. Conclusions

The stable structure of cobalt at room temperature is found to depend upon the grain size. When the grain size is very small, as in cobalt sponge, the stable structure is face-centred cubic, and this remains the stable structure up to a temperature of at least 600°C. When the grain size is large, as in annealed material in the form of a solid rod, the stable structure is close-packed hexagonal from room temperature up to about 450°C. Above this temperature the structure is face-centred cubic. When the grains are intermediate in size, as in fine unsieved filings taken from the rod, the two structures exist together in amounts depending upon the range and proportion of grain sizes in the specimen under consideration. In a specimen of filings quenched from above about 450°C it is possible to obtain a face-centred cubic structure with a lattice parameter considerably larger than that of the cubic lattice previously found; in fact during a long annealing period at 595°C it was possible to obtain records of the smaller cubic lattice ($a = 3{\cdot}5370$ kx) changing into a larger cubic lattice ($a = 3{\cdot}5540$ kx). It was the smaller cubic lattice that was encountered most often; the larger lattice was found only in specimens of cobalt filings quenched from temperatures between about 600°C and 840°C, and then it was found alone without a trace of the hexagonal lattice. It is probable that the material containing the larger cubic lattice after quenching is in a metastable state. We reject the possibility of contamination because (*a*) the samples are held at high temperature in evacuated enclosures, and (*b*) if there were contamination from the walls of the enclosure we should expect this to be more pronounced at the higher temperatures, but actually this effect was not observed at temperatures above about 840°C, and the lattice parameter then reverted to the smaller value it had at the lowest temperatures.

Different results were obtained according as the material was maintained and examined at elevated temperature or was quenched from elevated temperatures and examined at room temperature. It was found that all specimens of cobalt, whatever the grain size, annealed at about 1000°C and quenched in iced water, showed a mixture of the two structures, but in view of the foregoing statement it cannot be assumed that this also represents the state of the material *at* the high temperature. In fact the work of Newkirk and Geisler (1953) confirms the mixture of structures in material quenched from temperatures above 1000°C; their work further shows that when cobalt is maintained and examined at temperatures between 450°C and 1220°C it shows the presence of the face-centred cubic structure alone.

Combining these observations with the present results, it is concluded that the structure of cobalt sponge from room temperature to the highest temperature reached by the last-mentioned workers (1223°C) is face-centred cubic. But why

did the cobalt sponge specimens examined in the present investigation show a mixture of structures after quenching from temperatures of about 1000°C? The probable explanation is that at these high temperatures sintering took place, thus forming bigger grains and producing the condition necessary for the appearance of the hexagonal structure.

For large grains, such as are found in annealed wire specimens, the stable structure between room temperature and about 450°C is the hexagonal close-packed. Above this temperature the structure is face-centred cubic with lattice parameter at room temperature of 3·5370 kx.

The theory put forward by Edwards and Lipson (1943) satisfactorily explains the effects observed. They treat the problem as one of order–disorder. According to the theory the disorder term per gramme-atom is given by $-(\boldsymbol{R}/m)f(\alpha)$, where $\boldsymbol{R}$ is the gas constant, m the number of atoms in a plane of the crystal, α the probability of a fault in the structure and $f(\alpha)$ has a positive value. Thus the smaller the crystals the smaller is the number of atoms in the reflecting planes, and, therefore, the smaller (algebraically) is the value of the disorder term. Hence when the grains in the material are very small the material is in the ordered state and possesses a face-centred cubic structure. When the number of atoms in the crystal plane increases, the disorder term increases, and the material moves into the disordered state and assumes imperfect hexagonal close-packed structure. The experimental results recorded in this paper are explained on this view. The mixture of structures could also be explained on this theory if there were very small grains mixed large grains, as would undoubtedly be the case in samples of filings which had not been sieved.

The structures of cobalt and their lattice parameters are finally summarized in table 4.

Table 4. Cobalt Structures and their Lattice Parameters

Temp. range (°c)	Sponge	Filings	Rod
	(*a*) Material maintained at high temperature		
Room—*c*. 450	F.C.C. (a=3·5370 kx)	F.C.C.+C.P.H.	C.P.H. (a=2·5003 kx) (c/a=1·632)
c. 450–1220	F.C.C.	F.C.C.	F.C.C.
	(*b*) Quenched material		
Room—*c*.450	F.C.C.	F.C.C.+C.P.H.	C.P.H.
c. 450–*c*.900	F.C.C.	F.C.C. (changing from 3·5370 kx to 3·5540 kx)	F.C.C.+ C.P.H. (trace)
Above 900	F.C.C.+ C.P.H. (trace)	F.C.C. (small)+ C.P.H.	F.C.C.+ C.P.H.

F.C.C., face-centred cubic. C.P.H., close-packed hexagonal.

The curve showing the thermal expansion of cobalt sponge shows no abnormality; it is a smooth curve and, over the range of temperature investigated from room temperature to about 620°C, the thermal coefficient is a linear function of the temperature in accordance with the quadratic equation for the thermal expansion of the lattice derived from the observations. As the material is in a fine state of subdivision we are here dealing only with the face-centred cubic lattice. The thermal-expansion curve of cobalt when examined in bulk and containing grains of bigger size shows a discontinuity at about 450°C (Ellis and Greiner 1948). This

may be expected since the stable lattice in material of this nature, as already observed in the investigation, is close-packed hexagonal up to about 450°C, and it changes at this temperature to a face-centred cubic lattice, retaining this structure at higher temperatures up to 1220°C (Newkirk and Geisler 1953). The expansion of the hexagonal lattice has not been measured in this investigation; neither has the expansion of the larger cubic lattice found with quenched filings. It is of interest however to note that Marick has recorded values of the parameters of cobalt at elevated temperatures, which referred to the larger cubic lattice recorded by him. He found that the average coefficient of linear expansion of this lattice over the range 0°C to 800°C is 14×10^{-6}. If the curve of average coefficients drawn with values recorded in table 2 is extrapolated to 800°C it yields almost exactly the value found by Marick for this range of temperature, so that both cubic lattices show the same average expansion.

As the temperature is raised a transformation from hexagonal to cubic structure takes place at about 450°C; this is the only transformation, and it occurs only when the material contains grains large compared with those in the dust of cobalt sponge. The sponge alone does not show a transformation; it remains face-centred cubic from room temperature to 1220°C.

References

Cohen, M. U., 1935, *Rev. Sci. Instrum.*, **6**, 68.

Edwards, O. S., and Lipson, H., 1942, *Proc. Roy. Soc.* A, **180**, 268; 1943, *J. Inst. Met.*, **69**, 177.

Ellis, W. C., and Greiner, E. S., 1948, *Metals Handbook* (Cleveland Ohio: American Society for Metals), p. 1137.

Hendricks, S. B., Jefferson, M. E., and Shultz, J. F., 1930, *Z. Kristallogr.*, **73**, 376.

Hull, A. W., 1921, *Phys. Rev.*, **17**, 571.

Marick, L., 1936, *Phys. Rev.*, **49**, 831.

Neuburger, M. C., 1936, *Z. Kristallogr.*, **93**, 1.

Newkirk, J. B., and Geisler, A. H., 1953, *Acta Metallurgica*, **1**, 456.

Sekito, S., 1927, *Sci. Rep. Tôhoku Univ.*, **16**, 545.

Taylor, A., 1950, *J. Inst. Met.*, **77**, 585.

Wilson, A. J. C., 1942, *Proc. Roy. Soc.* A, **180**, 277.

Wyckoff, R. W. G., 1931, *The Structure of Crystals* (New York: Chemical Catalog Co., Inc.), p. 204.

Optical Studies of Ring Cracks on Glass

By S. TOLANSKY and V. R. HOWES
Royal Holloway College, University of London, Englefield Green, Surrey

MS. received 17*th February* 1954

Abstract. The surface distortions accompanying the familiar ring crack Hertzian cones formed on glass by local pressure are studied optically for four different optical glasses. The ring cracks are produced both with a rounded diamond and a steel ball indentor. The surface profiles are studied both by interferometry and with the light profile microscope. A frequent observation is the occurrence of multiple circular ring cracks surrounding a flat, slightly depressed disc, as a series of annular pile-ups of glass, usually not exceeding a height of 1500 Å. Surface profiles are illustrated for individual cracks on each of the glasses. The area of contact between glass and steel at the onset of a ring crack on a typical glass is revealed by an interferometric method. The crack appears to start at the rim of the circle of contact and sets in at a mean pressure of 14 000 kg cm^{-2}.

§ 1. Introduction

It has long been known that the pressure or impacting on a glass plate by a hard sphere leads to the formation of quite small circular ring cracks, frequently called Hertzian cones. These cones have often been studied and there is an appreciable literature dealing with them. In particular Preston (1921) and Bailey (1937) have made studies on these rings. It has been shown by sectioning that a circular ring crack has a conical split in depth. Bailey, without reporting specific measurements, proposes schematically a section as illustrated in figure 1 and also considers that severe strains arise due to lodgement of particles in the open cracks. Râman (1926) examined percussion figures interferometrically.

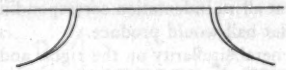


Figure 1.

The surface displacements which accompany the rings are in fact so small that they have not been previously analysed. We find that they can only be revealed by precision optical techniques. We have therefore undertaken an examination of the surface topography in the neighbourhood of ring cracks on several 'Chance' standard optical glasses. The surfaces have been examined by (*a*) multiple-beam Fizeau fringes (Tolansky 1948), (*b*) fringes of equal chromatic order (Tolansky 1948), (*c*) light-profile microscopy (Tolansky 1952).

The ring cracks were produced either by a steel or diamond ball indenter. The glasses used were small thick slabs of optically prepared glass of known specification and, in addition to their constitution classification, two distinct hardness numbers were available, namely (1) a pyramid indentation hardness and (2) a relative grinding resistance hardness.

§ 2. Experimental

The ring cracks were produced statically (pressure indent) either with a $\frac{1}{4}$ in. hardened steel ball, or a specially prepared diamond hemisphere (this was a cylinder with a hemispherical ball-end of diameter 0·78 mm), and dynamically with a falling steel ball. The largest ring crack made with the diamond had a diameter of 0·1 mm; hence for our purpose the ball-end cylinder can be treated as a diamond sphere. For the static cracks, loads in the range 2 to 4 kg were applied for 20 seconds. In the dynamic tests it was found convenient to drop a $\frac{1}{4}$ in. steel ball from a height of 80 cm.

After deformation, each specimen of glass was silvered by vacuum deposition following the practice now standardized for multiple-beam interferometry. The specimens were successively examined (*a*) by the microscope, (*b*) interferometrically, (*c*) with the light-profile.

Four glasses were studied having the following Chance's classifications: (i) Medium Barium Crown (MBC), (ii) Borosilicate Crown (BSC), (iii) Double Extra Dense Flint (DEDF), (iv) Light Flint (LF).

Details of surface topographies of static indentions will now be considered.

§ 3. Static Indentation

The interferogram in Plate I A is a multiple-beam Fizeau fringe picture, magnification × 160, taken with a fringe passing more or less diametrically over a ring crack. This crack was produced with a 4 kg load on the diamond ball. The following characteristics are evident. The ring crack is double, consisting of two approximately concentric rings of diameter 0·15 mm and 0·11 mm respectively. On approaching from the left to the outer ring from the undisturbed region (which we shall call zero level), the surface exhibits a smooth regular rise up to a height of 890 Å. On reaching the crack there is an abrupt drop, effectively very close to zero level, followed by a smooth rise, this time to 650 Å. Again on reaching a ring crack there is an abrupt drop almost to zero level. It is notable that the whole of the inner disc is near to zero level, the main disturbances being localized round the perimeter region.

There is in no sense at all an indentation corresponding to the Brinell indents on metals, which a similar ball would produce.

Although there is general similarity on the right- and on the left-hand sides, there is marked asymmetry in the amounts of the displacements on the two sides.

The Fizeau fringes are not linear profiles, but are mixed patterns determined both by the profile section and by the circular shape of the crack. For securing a diametrical topographical section the white-light fringes of equal chromatic order are distinctly advantageous. The remainder of the profile interferograms we show are therefore of this type. The distorted glass is silvered and matched against a silvered flat. Using white light, an image of the diameter of the ring crack is projected (with high magnification) on to the slit of a spectrograph. Each fringe formed is a true profile section, and in each pattern reproduced we include a small number of fringes to give the dispersion scale. For comparison with the Fizeau fringe pattern the interferograms have been placed with chromatic dispersion vertical so that the fringes act as topographical sectionings.

Although no two ring cracks made on any sample with the same load are identical, they are always fairly similar, and all show many characteristics in

common, which can be considered as typical for a glass. Hence the set of interferograms illustrated is only representative of many we have taken. An example of the similarity is shown by Plate I E and Plate I F, which both refer to the same glass.

Plate I B (×185) refers to MBC, with 4 kg, Plate I C (×185) is for BSC with 4 kg, Plate I D (×185) is for DEDF with 2kg, Plate I E (×185) is for LF with 4 kg, and Plate I F (×185) also for LF glass. The black line in I D is due to accidental scratching off of silver, and is to be disregarded.

The analysis of typical interferograms, drawn to scale, is shown in figure 2. The vertical scale is 400 times the horizontal scale.

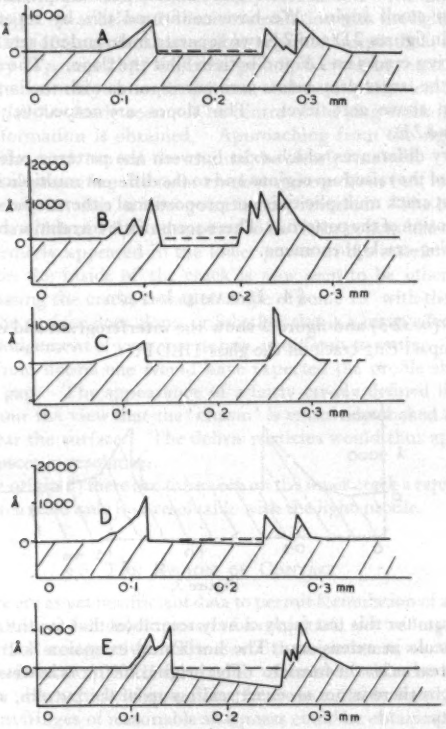


Figure 2.

In the following table are given a number of data derived from these diagrams: horizontal extension, maximum height of displaced material, and approximate volume of material displaced up above zero level. In addition we give the

Vickers indentation hardness and the grinding resistance hardness, both of these being supplied by the manufacturer; the grinding resistance is measured by rate of removal of material.

Glass	MBC	BSC	DEDF	LF
Horizontal extension (mm)	0·22	0·23	0·25	0·23
Vertical height (Å)	1000	1500	1500	1200
Displaced vol. ($cm^3 \times 10^{-9}$)	0·9	1·1	1·2	1·2
Vickers hardness,	615	600	420	460
Relative grinding hardness	1·88	2·07	1	1·95

A feature of considerable interest is the fact that the surface level within the ring crack is slightly below zero for MBC and BSC, the respective values being very small, 60 and 200 Å. The LF glass reveals not a simple displacement, but a slope of very small angle. We have confirmed this by repetition (Plate I F) and include in figures 2 D and 2 E two separate independent sets of measurements on separate ring cracks on LF and both exhibit the slope. The direction of slope is such that the larger drop below zero corresponds with the larger mass of glass displaced up above zero level. The slopes are respectively 20″ and 1″ for figures 2 D and 2 E.

Secondary differences which exist between the patterns refer primarily both to the slopes of the raised-up regions and to the different multiplicity of ring cracks. It is clear that crack multiplicity is not proportional either to the size of the height or to the extension of the pattern. There seems little correlation between hardness figures and ring-crack phenomena.

§ 4. Dynamic Impact

Plate I G (× 32·5) and figure 3 show the interferogram and interpretation for a dynamic impact ring crack on the glass DEDF.

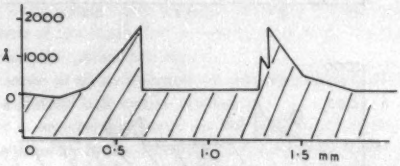


Figure 3.

The diagram for this test fairly closely resembles that for the static indent but on a bigger scale in extension. The horizontal extension is 1·2 mm, i.e. some 4·3 times as great as in the former. The height, 2000 Å, is however only 1·3 times as great. Thus there is not an exact scaling-up of the pattern, although there is close similarity.

§ 5. Examination with Light-Profile

It is familiar feature of multiple-beam interferometry that the high resolution in depth is achieved effectively at the expense of resolution in extension, and it is not usually possible to secure large magnification or resolution in extension simultaneously with the high resolution in depth. On the other hand, the light-profile,

when used with an oil immersion objective makes available magnifications up to $\times 2000$ both in depth and in extension simultaneously, and also readily resolves $0{\cdot}25\,\mu$ in each direction. Such powers in both directions simultaneously are of some considerable value.

Since high magnification in extension must accompany high magnification in depth, light-profile examination can only be applied in succession to different regions of the ring cracks, for the pattern extends too much for high magnification to encompass a whole ring. We show in Plate II A, at magnification $\times 850$ both in depth and in extension, a light profile crossing some of the multiple ring cracks on the dynamic test on DEDF. This has been selected because of the larger scale of the phenomena as compared with those obtained by static pressure. (The height displacements in the latter, whilst easily measurable with multiple-beam interferometry, are just on or below the limit of resolution with the light-profile).

In Plate II A three distinct cracks are visible, separated respectively in extension across the glass by $23\,\mu$ and $25\,\mu$, with the principal crack in the centre. The profile line passes horizontally across the pattern, and the displacement at the major crack, although small, is clearly seen. The centre of the ring crack is to the left. The following information is obtained. Approaching from the right, the displacement at the first crack can (on the original) be clearly resolved. The gradual smooth rise up on approaching the main crack is quite evident but, in addition to the former interferogram data, a new point is now revealed. The crack region, now resolved, has a width of $1{\cdot}7\,\mu$ and appears to be either conchoidal or filled with debris. What formerly appeared in the lower power of the interferogram to be an abrupt drop on the inside of the crack is now seen to be otherwise. The profile line, in crossing the crack, makes an angle of some 15° with the horizontal. Whether in fact the surface does slope, or whether this is a mean effective surface produced by the lodgement of very fine debris, is difficult to settle. With a wide open crack free from debris one would have expected the profile shadow to be missing over the gap. The appearance of a fairly crisply defined line over the 'chasm' does favour the view that the 'chasm' is effectively choked up with fine debris, at least near the surface. The debris particles would thus appear in size to be below microscopic resolution.

Finally (on the original) there are to be seen on the inner crack a repetition of the same effect, but on a scale only just resolvable with the light-profile.

§6. The Region of Contact

Although there are as yet insufficient data to permit formulation of a mechanism of the onset and ultimate full development of a ring crack, yet we have carried out one experiment which is suggestive. The aim of the experiment was to observe on the glass, by interferometry, both the region of contact with the ball and the surrounding area. To achieve this a glass plate was suitably lightly silvered so that multiple-beam fringes of reasonable sharpness could be obtained between the polished steel ball and the glass.

Using an inverted type microscope, the silvered glass was subjected to pressure from above with a $\frac{1}{8}$ in. steel ball. Simultaneously, from below, interference fringes formed between the glass and sphere were observed. Increasing loads were applied slowly and short intervals allowed between each load. At 3·6 kg a crack started from a point and rapidly spread almost into a complete circle.

A photograph of the superposed ring crack and fringe pattern is shown in Plate II B. There is a good deal of scattered light in the region of the crack, which photographs as black. The crack forms an incomplete circle with slightly spiral character. The picture shows fringes, but these only begin to appear outside the disc of contact between glass and steel ball. The edge of the disc of contact is a circle whose radius is about half a fringe-order separation less than the radius of the first (inner) fringe.

It will be seen that the ring crack begins almost exactly on this rim of contact. The crack as it developed did not follow this boundary but deviated spirally outwards so that on completion there are three light waves, say $1{\cdot}6\,\mu$ of air space, between the cracked glass and the metal sphere.

The irregular patches within the inner circle are to be disregarded since they represent regions where silver has been rubbed off the glass during the indenting process.

These observations clearly establish that the ring crack begins at one point on the rim of contact and then proceeds to develop effectively round this rim, although not exactly coincident with it, and distinctly moving outwards away from the region of contact. The region of contact itself is a perfect circle.

Since loads were slowly applied, the condition shown does in fact represent effectively the exact condition for onset of crack. The measured area of contact is $2{\cdot}6 \times 10^{-4}\,\text{cm}^2$, and as the crack developed at a load of 3·6 kg this corresponds to a stress of almost $14\,000\,\text{kg cm}^{-2}$, which is some 34 tons in^{-2}. The photograph in Plate II B was taken whilst the glass was under strain. The ring crack is therefore wide open and for this reason appears so very broad. On release of the strain the ring crack becomes very much narrower, in accordance with the usual observation.

We do find, however, that even the finest ring can be made much more visible if the glass be etched.

References

Bailey, J., 1937, *J. Amer. Ceram. Soc.*, **20**, 42.
Preston, F. W., 1921–2, *Trans. Opt. Soc.*, **23**, 141.
Raman, C. V., 1926. *J. Opt. Soc. Amer.*, **12**, 387.
Tolansky, S., 1948, *Multiple-Beam Interferometry* (Oxford: Clarendon Press); 1952, *Nature, Lond.*, **169**, 445.

Oriented Ring Cracks on Diamond

By S. TOLANSKY and A. HALPERIN *
Royal Holloway College, University of London, Englefield Green, Surrey

MS. received 17*th February* 1954

* Permanent address : Department of Physics, The Hebrew University, Jerusalem, Israel

Abstract. Two octahedral faces of a large very perfect diamond have been studied optically and interferometrically. On the surfaces have been found a hundred crater-like features each about 50 μ wide and a few hundred ångström units high. It is shown that these are the analogue on diamond of the long familiar ring crack Hertzian cones formed on glass subject to local pressure. It is possible that impact with other diamonds is responsible for the ring crack formation. The cracks on the diamond are crystallographically oriented and the profiles of some are measured interferometrically.

§ 1. Introduction

We have available a diamond with exceptionally fine surface quality and have submitted it to an intensive detailed optical examination. The techniques employed consist of a variety of multiple beam interference methods (Tolansky 1948), phase contrast microscopy (Zernike 1942) and the light-profile microscope (Tolansky 1952). The stone, from Sierra Leone, is a 'Portrait Stone', i.e. is an octahedron crystal in which a pair of parallel faces has developed, such that in effect the stone is a plane parallel hexagonal plate, some 1 square centimetre in area and 2 millimetres thick. The two large plane surfaces, natural octahedra, appear to casual observation to be of perfect mirror-like quality. The stone is of high gem quality and worth more than £350.

The general surface characteristics will be described elsewhere by Tolansky and Omar. At low magnification the interferograms of the surfaces studied reveal a number of triangular growth sheets, apart from which the general surface area is reasonably flat, say within 1 μ. Examination at higher magnification reveals the existence of an astonishing concentration of very shallow trigons. Omar (1953) estimates over a million on each face and they are extremely shallow, their depths, as shown by precision interferometry, being mostly less than 50 Å. Some of these will appear in illustrations discussed here.

Amongst the mass of surface material under study curious features have emerged and this is the object of the present report. We have found no less than one hundred similar features which have a crater-like topography and it is the aim of this note to show that these very closely resemble the Hertzian (Hertz 1896) ring cracks on glass described by Tolansky and Howes (1954).

§2. Description of Features

We select a few typical examples an analysis of which throws light on their origin. Most are of about the same size, extending over only 50 microns and requiring therefore high power microscopes to show up their details. On all pictures the shallow trigons are much in evidence. Consider first Plate I (*a*).

The first characteristic of the central pattern of Plate I (*a*) is the hexagonal shaped outline and it can be seen from the orientations of the trigon sides that the hexagon is strictly crystallographically oriented. There are trigons both within and without the ring formation of Plate I (*a*). There are two other minor ring features in the figure, one at the lower right, the other at the upper left. The magnification here is $\times 600$ and it is significant that each of the three features appears to sit directly on a horizontal line, which is in fact a growth sheet edge.

Plate I (*b*) (magnification $\times 800$) with its two major details presents features which are revealing. The most notable characteristic shown in the upper figure is that the growth trigons cross the ring boundaries in strict straight line continuation. Since these shallow trigons are growth phenomena this would appear to establish conclusively that the ring markings are *not* due to growth. Plate I (*b*) is the best contrast obtainable by our phase contrast equipment. Plate I (*c*) ($\times 600$) is an interference contrast picture taken with a thin film technique and shows quite clearly that the lower ring formation has inner shallow steps.

The topography of these features is clearly established by Plate I (*d*) which is a high magnification multiple-beam Fizeau fringe picture ($\times 225$) taken diametrically across the two features of Plate I (*b*). The definition is exceptionally good when the high magnification is taken into account, fringes are shown for $\lambda = 5461$ Å. The fringes correspond to diametral sections running across the two features in a S.W.–N.E. direction, the two distorted fringes being those passing over the features. It will be seen that a fringe, on approaching a feature from outside rises smoothly up the side of the ring feature, and then drops abruptly, to such a position that the major region within the ring formation is effectively at the same level as outside.

There is complete identity with the fringe patterns shown by Tolansky and Howes (1954) *on the circular ring cracks in indented glass*, and we shall give later additional support to show that this is the most probable explanation of these phenomena on the diamond. It is of incidental interest that the region surrounding the ring features is in Plate I (*b*), free from trigons and this is confirmed by the straightness of the fringes.

Plate I (*e*) shows a further characteristic example (magnification $\times 287$) and finally, Plate I (*f*) and I (*g*) are other examples ($\times 400$) in which the ring features clearly cut across growth trigons, both large and small.

§3. Discussion

We shall now discuss the origin of the ring markings. The major characteristics of the ring features appear to be as follows :

(i) All appear as small crater-like elevations which are usually a few hundred ångström units above the general surrounding level, as measured by interferometry.

(ii) The inner main region is effectively at the same level as the outer region.

(iii) The approach to the ring crater from outside is gradual and smooth, the descent inside is abrupt, and may be stepped.

(iv) The patterns tend to be approximately linear configurations with sides frequently oriented in the close-packed directions of the crystal.

(v) Most noticeable is the manner in which the growth trigons and growth sheets carry on right through the markings, despite the elevation differences.

(vi) The margins of any successive edges tend to reflect any irregularities which may appear in the first edge.

We shall consider various possible explanations of the origin of the ring markings and believe that we have established their nature with some certainty.

From the above list of characteristics it seems quite clear that the rings are not due to etch or solution figures. Apart from this, there are no traces at all on the rest of the crystal of the familiar triangular pits of etching, which we always have found by direct experimental etching to be oppositely oriented to the growth trigons and therefore easily identifiable. One must consider too whether the markings are imprints left by inclusions which have been released with time. This can be rejected since it neither accounts for the observed crystallographic orientation nor can it explain the apparent penetration into the rings of the growth trigons. This latter fact indeed entirely disposes of the possibility of these features being due to growth. They must have appeared after growth ceased.

It is our view that these ring features are the analogue in diamond of the ring cracks described for glass by Tolansky and Howes. As pointed out there, it has been known for over 50 years that small ring cracks called Hertzian cones, are formed by local pressure on glass by a small ball. There is almost exact correspondence with the interferometric profile of the glass ring cracks and the diamond ring features. In both cases a smooth approach from outside, rising to a crater some hundreds of ångström units high, and a sharp drop to a level practically the same as outside the feature, with frequent appearance of multiple edges.

Thus we conclude that the ring features are the Hertzian cones formed by external indentation, after the growth of the diamond has ceased. This is a very satisfactory explanation of all the topographical features, accounting for example for the continuation of growth details through the ring features.

The principal difference between the glass and diamond markings is the fact that the former are circular and the latter crystallographically oriented. This also is explicable for clearly cracking is likely to be preferred along crystallographic cleavage directions.

Regarding the origin of these ring cracks, there would appear to be two alternatives, namely :

(*a*) During its history the crystal may have suffered contact from other diamonds. Repeated impact from one or more diamonds would have caused the hundred or so ring cracks observed. Or again pressure against a magma containing numerous microcrystals could be responsible. The very planeness of the portrait stone exposes its whole area to such impacts even from microcrystals. In this connection it is of interest to note that some evidence has recently been reported that diamonds can produce indents on each other when sharp points are impacted on to plane surfaces. This owes its existence to the differential hardness of diamond in different directions.

We have not yet established whether any material other than diamond could cause indents on diamond.

(*b*) An alternative possibility, but of much less weight than the above, is that internal strains have been responsible. In possible support of this is the

observed tendency for ring features to be located on growth sheet edges. We do not attribute much weight to this explanation, because in fact the crystal is of very fine quality and free from strains. It is always possible that the surface has received many more impacts than there are ring cracks and that an impact near a growth ledge is more favoured for producing a ring crack than an impact elsewhere. This would result in a statistic favouring the appearance of ring cracks near growth ledges.

It is our intention in the future to undertake experimental work with other diamonds in the hope of being able to induce ring cracks in the laboratory by indenting with prepared diamonds.* We have observed ring cracks on many diamond octahedral faces in addition to that discussed here.

Acknowledgments

One of us (A. H.) wishes to express his thanks to UNESCO for the award of a scholarship, and to the Hebrew University, Jerusalem, for granting study leave.

References

Hertz, H., 1896, *Papers*, translated by Jones and Schott, Chap. V, p. 176 (London : Macmillan).

Omar, M., 1953, *Thesis*, University of London.

Tolansky, S., 1948, *Multiple-Beam Interferometry of Surfaces and Films* (Oxford : University Press) ; 1952, *Nature, Lond.*, **169**, 445; 1953, *Research, Lond.*, **6**, 7.

Tolansky, S., and Howes, V. R., 1954, *Proc. Phys. Soc.* B, **67**, 467.

Zernike, F., 1942, *Physica*, **9**, 686.

* *Note added in proof.* This we have now succeeded in doing, on a variety of crystal faces.

Variations in Normal Colour Vision in Relation to Practical Colour Matching

By F. L. WARBURTON

Wool Industries Research Association, Torridon, Headingley, Leeds 6

MS. received 4th December 1953; read before the Colour Group 6th December 1950

Abstract. The influence of variations in colour vision on matching a dyed pattern with one of a range of dichroic series has been observed for 250 normal observers. The results have been compared with the variations produced by a change in lighting for a single observer. The results have been related to the data of Wright on the variations in colour vision attributable to differences in pigmentation with which they are in close accord. The number of observers was sufficient for a statistically significant change in colour vision with age to be detected. This change corresponds to increased pigmentation and confirms the commonly held view that older people see redder than young. The bearing of the results on the selection of colour workers is discussed.

§1. Introduction

The *Report on Defective Colour Vision in Industry* published by the Colour Group of the Physical Society a few years ago stated that in certain industries a small proportion of the personnel required exceptionally good colour vision. Such workers included colour passers in dye houses, etc. In these cases the standard of colour matching required is very high and where coloured surfaces having different spectral reflectance curves have to be matched, it is not sufficient merely to eliminate defective colour vision but the variations in normal colour vision have also to be considered. The most important of these variations is that which is usually attributed to absorption in the eye mainly by the macular pigment. Since these differences are such as could be produced by putting a yellow filter in front of the eye, their effect in practical colour matching is very similar to that of varying the colour temperature of the light source. The only numerical data on these variations so far available are those due to W. D. Wright (1946, p. 133), who calculated the chromaticities of different stimuli for a number of observers with the aid of a convention that showed these differences in a simple way; the normal colorimetric conventions are designed to reduce the influence of such variations.

The work of Wright as well as the practical experience of dyers, suggests that the influence of these differences in practical colour matching is considerable, corresponding to that of changing the light source from a photoflood lamp to North daylight, i.e. over a colour temperature range of 3500°K to 7500°K. The occasion of an exhibition of artificial daylight lamps* was therefore taken as an opportunity to make a survey of such variations as it affected the matching of patterns dyed from different recipes. Altogether 250 observers were examined

* *Light and Lighting*, August 1950, **43**, 298.

(mainly connected with the wool and related industries) and the variations found corresponded with those expected from Wright's data. The number tested was sufficient to enable a correlation to be obtained between age and amount of pigmentation, thereby confirming the general belief that pigmentation increases with age.

§2. Previous Data

Although the existence of considerable variation in colour vision of a type which could be attributed to variations in macular pigmentation is well known, the only convenient numerical data are those obtained by Wright (1946, p. 133) as part of his data which are incorporated in the specification of the C.I.E. standard observer. The method which he used to eliminate the effect of such variations on the chromaticity coordinates for the spectral colours by basing his units on two matches on spectral colours was particularly useful in showing these differences between observers as a variation in the chromaticity of the standard white. The principle of the method has been fully discussed by Wright and need not be detailed here. Wright's results are shown in figure 1, and as a scale for the magnitude of this variation the black-body locus for the standard observer on

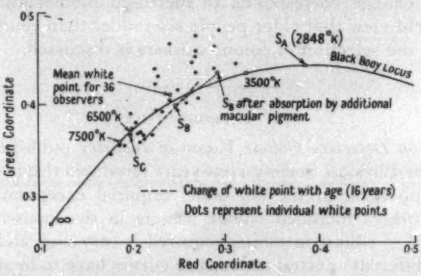


Figure 1. Chromaticity diagram in W.D.W. units showing the white points for real observers and the black body locus and coordinates of important illuminants for the standard observer.

the same system has also been plotted. Additional data show the change in chromaticity of the N.P.L. white with age as given by Wright, and also the chromaticity of C.I.E. S_B source after passing through a medium having average absorption values of the macular pigment quoted by Parsons (1924). This was calculated on XYZ system by the ten selected ordinate method and transformed to the W.D.W. system. The reverse calculation has also been made which shows that the chromaticity of S_B after neutralizing the absorption in the macula is similar to that for S_C.

Co-ordination of all these data shows the magnitude of the variations in the so-called macular pigmentation and also shows that the absorption of pigment taken from the eyes of deceased persons is consistent with the traditional attribution of these variations to macular pigmentation. Comparison of the observed shift in chromaticity of N.P.L. white with age with that calculated suggests that the increase in ocular absorption with age may be accompanied by a change in the hue of the absorbing media. The change in hue may be due either to a change in the absorption characteristics of the macular pigment or the additional

absorption may be due to other causes such as yellowing of the lens. Similar calculations have been made by Von Schelling (1950) using different data on the absorption of the macula. This author has also calculated the possible shifts in chromaticity with change in illumination for all possible spectral energy distributions which could produce any given initial chromaticity.

Although it is widely believed that the amount of absorption in the eye increases with age the only definite evidence is that of change of the chromaticity of the N.P.L. white for W. D. Wright (1946, p. 153) with increasing age. The range of pigmentation of different individuals of the same age is too great to allow a definite conclusion to be drawn from investigations with a small number of observers of different ages. In fact no such conclusion could definitely be drawn from the 36 points shown in figure 1, as the point showing the greatest pigmentation was given by a student of about 20 years of age (Wright 1946, p. 153).

Other data supporting the view that there is a change of colour vision with age of this type have been published (e.g. Knowland 1945). However, the number of subjects tested in each case has been insufficient to establish the correlation with age.

§3. The Present Investigation

The investigation described in this paper was subsidiary to an exhibition of daylight lamps intended for accurate colour matching. One of the objects of the survey was to obtain evidence in a form easily appreciated by the practical dyers that the variations in the colour matching characteristics of people who in no sense had defective colour vision had as great an influence on practical colour matching as very large variations in the colour temperature of the light source used. It was also hoped to discover whether smaller variations were found among dyers than among the general population as a result of selective recruitment.

The above conditions largely determined the nature of the test used. These implied that the test should be simple and related as closely as possible to the ordinary work of the dyer. The commercial test which best satisfies these conditions is the Glen Colorule made by Sydney Blumenthalls, U.S.A., but the author was unaware of the existence of this device until a few days before the exhibition, and it was doubtful whether spectrophotometric data were available for the test patterns. The test also suffers from the disadvantage that the patterns used are rather small. However, a similar but simpler test had been devised some years ago by the Bradford Dyers Association although it had not been made on a commercial scale. This test consists of a slightly dichroic standard pattern which is to be matched against one of a numbered range of dichroic patterns dyed with different dyes from those used for the standard, the position of the match varying with the observer and the illumination under which it is used. The patterns in the numbered range were each about $2\frac{1}{2}$ inches long by 1 inch wide and varied in hue from a dull green at No. 1 to a reddish brown at No. 13. All these patterns were redder in artificial light than daylight whereas the standard was greener. This set of patterns was made available to the author, and spectrophotometric data were later obtained on the Hardy recording spectrophotometer. The patterns were illuminated by the S_B source and visitors were asked to fill in on a card the number of the pattern in the range which matched the standard. The card also contained spaces for additional information such as the age and occupation of observer.

It was impossible to make a preliminary test for defective colour vision but it is unlikely that any such person would be successfully engaged in work closely related with dyeing at any rate without being aware of the defect. The comments made on the difference between the standard and the nearest pattern in the range were also such as to eliminate the more serious cases of anomalous trichromats since from the nature of the test it is unlikely that they would find a satisfactory match with any of the range of patterns. This was in fact confirmed by a card returned by a visitor who knew that he had defective colour vision. A few other cards were returned by visitors who were known to have defective colour vision. Two of these, of whom one was a deuteranope and the other a protanope, had been tested by the Ishihara cards. The difference between their cards was remarkable and will be referred to later. It may, therefore, be assumed that if any cases of defective colour vision are included in this survey they will be too few to influence the results.

§4. Results of the Survey

The chromaticities, for illuminant B, of the patterns used were calculated from curves obtained on the Hardy spectrophotometer using ordinates at 5 mμ intervals. The curves for patterns 5, 8 and 11 are shown in figure 2. These were plotted on a uniform chromaticity basis. For this purpose a method,

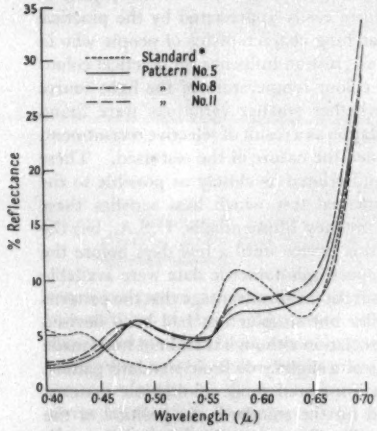


Figure 2. Hardy spectrophotometer curves for selected patterns.

* Slightly fluorescent; the total amount of light reflected in blue will be greater than shown.

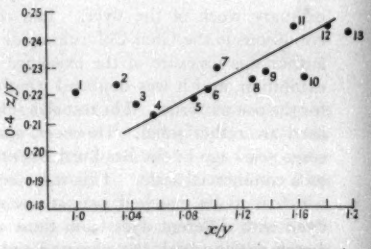


Figure 3. Chromaticities of patterns used plotted on a uniform chromaticity diagram.

based on one suggested by E. Q. Adams (1942), in which x/y is plotted against $0{\cdot}4z/y$ was used, as being the simplest for computational purposes. For a short range of colours lying approximately on a straight line in the chromaticity chart there is very little to choose between different systems; furthermore, the author is of the opinion that an equal brightness chart gives as close an approximation to uniform chromaticity as any except for the saturated blue and blue green

regions. The results are shown in figure 3. It will be noticed that the patterns numbered from 4 to 12 with the exception of number 10 lie very close to a straight line. As all the matches obtained lay between numbers 5 and 11 the best straight line was drawn through these points excluding number 10 which was noticed visually to be out of step. It would be expected that an actual observer using the dyes with which the range was dyed would match the standard with a combination giving a colour lying along the line, and that the distance along this line between the points representing the matches made by any two observers would be proportional to the visual difference between these matches for the standard observer. The analysis of the cards was, therefore, made on the basis of the distance along this line from the point representing pattern 5.

The spectrophotometer curve for the standard pattern is also shown in figure 2. This pattern was however slightly fluorescent so that the curve does not give a true picture of the light reflected from the surface under any particular illumination and the chromaticity cannot be calculated from this curve. It does however indicate the dichroic nature of the pattern and why it changes in colour in a different way from the numbered range when the illumination is changed. This conclusion is confirmed by subsequent spectral reflectance measurements made on the G.E.C. Photoelectric Eye and Physical Colorimeter in which the dispersing system is placed after the sample. The general shape of the curve is not affected by this procedure but the relative reflectance increases progressively towards the blue end. The chromaticity of this pattern for the standard observer was also determined by this instrument; at the same time measurements were also made on such of the range of patterns as were available in larger pieces. Owing to the low luminance factor of the patterns the accuracy attainable by this instrument and the recording spectrophotometer was insufficient for the two sets of data to be combined. It therefore proved impossible to determine the match which would be found by the standard observer. The effect of change of illumination on the position of match was therefore determined for the author's eye by observing the patterns under the light of illuminants A and B and an under-run Macbeth colour matching lamp giving an illuminant close to S_C. The matches found in this way were as follows: S_A, No. 3; S_B, No. 8; Macbeth, No. 12.

Figure 4 shows the result for 247 observers plotted in the form of a block diagram the abscissae being plotted on a uniform colour step scale. The ordinates represent the number of observers finding the nearest match with each pattern or between adjacent pairs. As might be expected there is a tendency for the height of the ordinate to be a little lower for the intermediate positions as these probably cover a smaller range of positions for the exact match. In obtaining the position of the mean match the number of observers in each block was multiplied by the distance of the mid-point in each block from the point representing pattern 5, and the total divided by the number of observers. The mean position found in this way falls about mid way between the points representing the chromaticities of patterns 8 and 9, and the 95% confidence limits, i.e. the range within which the true position is likely to be found, covers the width of the block representing a match between 8 and 9.

The range of matches is from pattern No. 5 to pattern No. 11. In comparing this with the variation of the position of the best match with change in illuminant for the author it should be noted that the latter finds a match under S_B which corresponds to slightly greater pigmentation than the average. Even allowing

for this it is seen that the range of matches for different observers does not differ much from that which would be found for the standard observer with illuminants varying from S_C to about halfway between S_B and S_A. This corresponds closely with what would be expected if the scatter of the white points in figure 1 was due to absorption in a medium which is an approximate colour temperature corrector. The distribution of matches is also not dissimilar to that of the white points in figure 1. The relatively large number of matches obtained with pattern 11

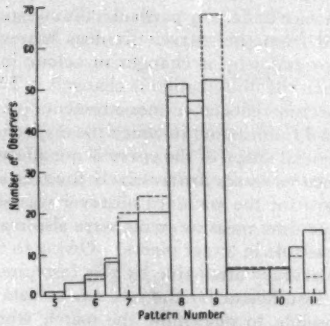


Figure 4.

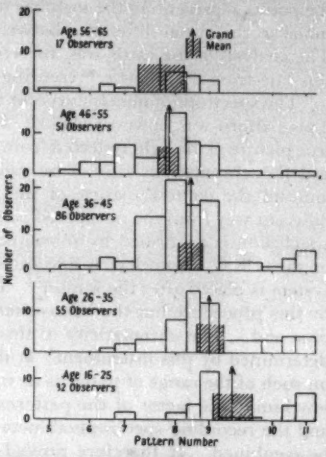


Figure 5.

corresponds to the bunch of points round the S_C point in figure 1, and the drop between 11 and 9 although probably due, in part, to the fact that the chromaticity of pattern 10 was out of line, is not dissimilar to the region of fewer points between this group and the much larger group around the mean position of the N.P.L. white point.

There is no doubt that the range of matches found is unaffected by the possibility that the survey includes a few colour defective observers, since many of the matches with pattern 11 were made by females. Some of these have also been tested as a further check. At the other extreme it is almost certain that the observer who found a match with pattern 5 has been tested at some time or other and from the position he holds it is certain that if he had defective colour vision it could not have escaped detection.

Figure 5 gives the data for each of five age groups plotted in a similar manner to figure 4. Each group covers a period of 10 years. The position of the mean match for each group together with the 95% confidence limits ($\pm$ twice the standard error of the mean) is also shown. Examination of the figure shows that the change in position of the mean is definitely significant for alternate age groups but not always so for adjacent groups. The analysis by age groups thus provides concrete evidence that there is a definite increase in the degree of pigmentation with age. An attempt was also made to see whether there was any appreciable difference in distribution in pigmentation between dyers and the general public.

The view has occasionally been expressed that when a young dyer is appointed he is often selected as having similar colour vision to the senior dyer. The latter often has a large amount of pigmentation, with the result that when the second man in his turn appoints an assistant, a still more pigmented observer is chosen.

The cards were, therefore, sorted so as to select those whose occupation suggested that their duties included practical colour matching or shade passing. There was naturally a certain amount of uncertainty in making the selection as the description of the observer's occupation was not always sufficiently precise. (It was impossible to provide continuous supervision of the test.) About a quarter of the total number of observers fell into this category and the results are shown in figure 4 by dotted lines on a scale four times as great as that for the general survey. These show no significant differences from the general survey except that there are relatively fewer cases of extreme over or under pigmentation. This would imply that the biased selection of dyers suggested above is not of frequent occurrence, partly because many senior dyers, particularly with large firms, are aware of the change with age. Supporting evidence for the non-biased selection is found in the fact that the whole range of pigmentation as found among dyers occurred among the personnel of one firm.

§ 5. General Discussion of the Results

Comparison of figures 1 and 4 shows that the influence on practical colour matching of the variations in retinal pigmentation found by Wright is as great as that of the change in light source required to produce a chromaticity difference on the W.D.W. system as that shown for different observers for the N.P.L. white. The difference is seen to be considerable, and thus standardization of the colour of light sources for colour matching is of very little value unless the observer is also standardized at somewhere near the average pigmentation value. A further deduction is that less stringent limits are required on the colour temperature of a source having a black-body colour than on departures from a black-body energy distribution.

The analysis also shows that as a long-term policy such standardization of the observer will not give rise to any serious difficulties since 60% (150 out of 247) found matches ranging between patterns 8 and 9. Immediate standardization would, of course, cause hardship to existing personnel who did not comply with the standard.

The demonstration of the importance of pigmentation variation raises the question of an appropriate test when selecting colour workers. The most satisfactory absolute test would be some form of vector colorimeter which would enable the white point for the subject to be plotted on the chromaticity chart of figure 1. Such an instrument would, however, require proper laboratory facilities for its operation. For general use an improved form of the test used in the present investigation would appear to be a suitable one. Such a test would require standardizing in the first instance against more fundamental data and subsequent manufacture should be under close colorimetric control. Should a new determination of the standard observer be made the easiest way of making the initial calibration would be to apply the test to each observer taking part in the determination. The only existing device suitable for this purpose is the Glen Colorule but it should not be impossible to produce a similar device in this country.

Connected with this consideration is the selection of observers for any such determination of the standard observer; owing to the variation in average pigmentation with age it is important that the personnel used in the determination of the standard observer should be as representative as possible of all adult age groups. Failing this it would appear from the analysis of the survey that a distribution of ages symmetrically grouped around 40 years would give a close approximation to the average at least for large field size.

§6. Results with Observers having Defective Colour Vision

As mentioned above two of the cards were returned by observers who had been tested elsewhere with the Ishihara cards, one being recorded as a protanope and the other as a deuteranope. The former found a match between patterns 9 and 10 thus falling well within the range of normal observers. The latter found a match between the reference pattern and pattern No. 3. Another observer known to himself to be an anomalous trichromat found no match as all the patterns were too red.

These results may be compared with anomaloscope tests where it is found that deuteranomalous observers make matches quite distinct from the range of normal observers whereas protanomalous observers are more difficult to distinguish. It has also been suggested (Wright 1946, p. 304) that deuteranopes would differ less in this respect from the normal than would deuteranomalous observers.

Acknowledgments

The author's thanks are due to Dr. A. B. D. Cassie, Director of Research, for his interest in this work and to the Council of the Wool Industries Research Association for permission to present this paper. He is also greatly indebted to Mr. E. Wilson of the Bradford Dyers Association for the loan of patterns used for the test, to Dr. E. Waters and the staff of I.C.I. Dyestuffs laboratory for the spectrophotometric data and to Mr. G. T. Winch and Miss B. M. Young for measurements on the photoelectric eye. Thanks are also due to the staff of the Statistics Department of W.I.R.A. for many of the colorimetric calculations.

APPENDIX

Since the derivation of the transformation equations from the C.I.E. XYZ system to the W.D.W. system is somewhat laborious, these are given here in case other workers should desire to make use of them. They are:

$$X = 0{\cdot}817708R - 0{\cdot}504076G + 0{\cdot}018598B$$
$$Y = 0{\cdot}150264R + 1{\cdot}335662G - 0{\cdot}049313B$$
$$Z = 0{\cdot}137102R + 0{\cdot}039837G + 0{\cdot}935642B$$

References

Adams, E. Q., 1942, *J. Opt. Soc. Amer.*, **32**, 168.
Knowland, E. P., 1945, *Amer. Dyest. Rep.*, **34**, 308.
Parsons, J. H., 1924, *Colour Vision*, 2nd Edn (Cambridge: University Press), p. 44.
Von Schelling, H., 1950, *J. Opt. Soc. Amer.*, **42**, 419.
Wright, W. D., 1946, *Researches in Normal and Defective Colour Vision* (London: Henry Kimpton).

Theory of the Positive Column

By G. ECKER*
Physics Department, Queen's University, Belfast

Communicated by K. G. Emeléus; MS. received 12th February 1954

Abstract. Experiments have shown that when the current through a positive column in mercury vapour falls below a certain limit, the longitudinal field does not remain constant, but increases with decreasing current, at first slowly and later more rapidly. This is due principally to failure of the discharge to establish a plasma with approximately equal concentrations of ions and electrons. The properties of the positive column have been formulated without making any initial assumption about the concentrations. A complicated system of differential equations with boundary conditions results, which determines the concentration of the charged particles, and hence all the other properties of the discharge. Schottky's theory of the quasineutral column appears as a limiting case. To obtain a preliminary solution of the equations, which need normally mechanical integration, the assumption has been made that, to a first approximation, the concentrations are in a constant ratio. The equations can then be solved, and in particular expressions obtained for the current *vs.* longitudinal field characteristic curve for the column, and for the onset conditions for the special properties of the low-current form of the column. Fairly good agreement has been found between the numerical predictions of the theory and the results of experiment. It has also been found that with decreasing current density the field approaches a limiting value which is obtained from the usual Schottky formula simply by replacing the ambipolar diffusion coefficient by the electron diffusion coefficient.

§ 1. Introduction

Mierdel and Schmalenberg (1936) have made measurements of the longitudinal field E_z in cylindrical positive columns in mercury vapour at pressures such that the free paths were small compared with the tube radius R. Their results, which are reproduced in the figure, show that for very small currents E_z increases rapidly as the current density j decreases. This deviation from the Schottky theory (1924), which, on the assumption of quasineutrality predicts that E_z should be independent of j, is too great to be explained by secondary effects such as cumulative ionization or change of gas temperature and density (Mierdel and Schmalenberg 1936). It has been suspected that in this form of positive column, which will be referred to as 'subnormal', the concept of quasineutrality fails (Mierdel and Schmalenberg 1936). The main aim of the present paper is to formulate a general theory for the positive column without Schottky's restriction that there are plasma conditions (quasineutrality), and to obtain an approximate solution of the equations for the column which will account for the falling current–voltage characteristic curve.

* On leave of absence from the Institute of Theoretical Physics, University of Bonn.

We adopt the following picture of conditions in the discharge. In the column there are electrons and ions whose motion in the longitudinal electric field transports the tube current. Simultaneously the charged particles move towards the walls under the combined influence of the concentration gradient and ambipolar field. They neutralize one another on the walls. Since recombination in the gas does not occur to any appreciable extent on account of the small concentrations, the loss of charged particles to the walls must be compensated by fresh production of ions and electrons by electron collision with gas molecules. The fundamental property of the subnormal column is that the concentrations of electrons n_e and ions n_i are small. This has the consequence that the difference $n_i - n_e$ needed to produce the space-charge responsible for the ambipolar radial field is not small compared with n_e and n_i, as it is in a plasma, and makes the ordinary Schottky theory inapplicable, although the latter must be the limiting case of the general theory when the current becomes high.

§ 2. Development of the General Theory

The differential equations governing the density distributions of the charged particles must be developed *ab initio*, as the simplifications of Schottky's theory are not applicable in the more general treatment. The basic equations are: (*a*) Maxwell's equations, which take a simple form, as we deal with a stationary state, and the magnetic field of the current can be neglected; (*b*) the balance equations for electrons and ions; (*c*) Boltzmann's transport equation for determining the distribution functions of the ions and electrons.

The equations (*a*), (*b*) and (*c*), together with the boundary conditions, define uniquely the properties of the column.

The first step towards the solution consists in finding from (*c*) the transport equation of the ions and electrons. As is well known, the movement of the charged particles is describable in terms of drift and diffusion, if n_e and n_i are small compared with the concentration of uncharged particles, and grad $n_{e,i}$ and $\mathbf{E}$ are not very large or strongly dependent on position. The corresponding mobility and diffusion coefficients b and D depend only on $\mathbf{E}$ and pressure p.

We can then formulate (*a*), (*b*) and (*c*) as follows:

$$\operatorname{curl} \mathbf{E} = 0 \qquad \ldots\ldots(1\,a)$$

$$\operatorname{div} \mathbf{E} = e(n_i - n_e)\epsilon_0^{-1} \qquad \ldots\ldots(1\,b)$$

$$\operatorname{div}(\mathbf{j}_i + \mathbf{j}_e) = 0 \qquad \ldots\ldots(1\,c)$$

$$\operatorname{div} \mathbf{j}_e = -en_e w \qquad \ldots\ldots(1\,d)$$

$$\mathbf{j}_e = eb_e n_e \mathbf{E} + eD_e \operatorname{grad} n_e \qquad \ldots\ldots(1\,e)$$

$$\mathbf{j}_i = eb_i n_i \mathbf{E} - eD_i \operatorname{grad} n_i \qquad \ldots\ldots(1\,f)$$

where ϵ_0 is the dielectric constant of vacuum, $\mathbf{j}_i$ and $\mathbf{j}_e$ the positive ion and electron current densities respectively, and w the ionization rate measuring the average number of ionizing collisions made by an electron per unit time.

In a uniform cylindrical column, symmetrical about the axis, (1 *a*) reduces to

$$E_z = \text{const.} \qquad \ldots\ldots(2)$$

and (1 *c*) becomes

$$\frac{d}{dr} r(j_{i,r}+j_{e,r})=0 \qquad \ldots\ldots(3)$$

where $j_{i,r}$ and $j_{e,r}$ are the radial components of $\mathbf{j}_i$ and $\mathbf{j}_e$ respectively.

From (3)

$$j_{i,r}+j_{e,r}=\text{const}/r \qquad \ldots\ldots(4)$$

and since neither $j_{i,r}$ nor $j_{e,r}$ can become infinite for $r=0$, the constant must be zero, and

$$j_{e,r}=-j_{i,r}. \qquad \ldots\ldots(5)$$

The relation (5) now allows the radial component of electric field E_r to be determined. Inserting the transport equations (1 *e*) and (1 *f*) in (5) gives

$$E_r=\left\{D_i\frac{dn_i}{dr}-D_e\frac{dn_e}{dr}\right\}\Big/\{n_ib_i+n_eb_e\}. \qquad \ldots\ldots(6)$$

The current densities $j_{e,r}$ and $j_{i,r}$ are found from (1 *e*) and (6) to be

$$j_{e,r}=-j_{i,r}=e\left\{n_eb_eD_i\frac{dn_i}{dr}+n_ib_iD_e\frac{dn_e}{dr}\right\}\Big/\{n_eb_e+n_ib_i\}. \qquad \ldots\ldots(7)$$

Introducing (6) and (7) in (1 *d*) and (1 *b*), we obtain

$$\frac{1}{r}\frac{d}{dr}\left\{r\frac{n_eb_eD_i\,dn_i/dr+n_ib_iD_e\,dn_e/dr}{n_eb_e+n_ib_i}\right\}=-n_ew \qquad \ldots\ldots(8\,a)$$

and

$$\frac{1}{r}\frac{d}{dr}\left\{r\frac{D_i\,dn_i/dr-D_e\,dn_e/dr}{n_eb_e+n_ib_i}\right\}=e(n_i-n_e)\epsilon_0^{-1}. \qquad \ldots\ldots(8\,b)$$

Equations (8 *a*) and (8 *b*) together with the boundary conditions

$$\frac{dn_{i,e}}{dr}=0, \qquad r=0 \qquad \ldots\ldots(8\,c)$$

and

$$n_{i,e}=0, \qquad r=R \qquad \ldots\ldots(8\,d)$$

determine n_i and n_e as functions of r.

When n_i and n_e are known, all the remaining properties of the discharge are determined. In particular, the current–voltage characteristic of the tube is given from (1 *e*) and (1 *f*) by

$$I=\int_0^R e(b_en_e+b_in_i)E_z2\pi r\,dr. \qquad \ldots\ldots(9)$$

The whole problem reduces therefore to the solution of equations (8 *a–d*), which transform to the equations of quasineutral ambipolar diffusion theory when $n_e\simeq n_i$.

§ 3. Approximate Solution

Evidently (8) is a complicated system of equations whose solution will necessitate mechanical integration. The analytical difficulties are enhanced by lack of experimental knowledge of the values of b, D and w. An approximate solution may be obtained by noticing that the radial variations of n_i and n_e must be similar. In the quasineutral case where $n_i\simeq n_e$ this is trivial. However, even when the last condition is not satisfied, both distributions must have a maximum at the axis, and diminish steadily to $n_{i,e}=0$ at the walls. These

restrictions do not give much latitude in choice of the forms of the curves for $n_{i,e}(r)$. We accordingly base our approximation on the assumption that the shapes of the curves, but not their absolute values, are similar, and that therefore

$$n_i = \gamma n_e \qquad \ldots\ldots (10)$$

where γ is a constant. The problem then reduces to the determination of such values of n_e and γ as give the best approximation to the exact expressions for $n_i(r)$ and $n_e(r)$. For this purpose we start from equations (8 *a–d*), noting that (10) is not in conflict with the boundary conditions. Introducing (10) in (8),

$$\frac{1}{r}\frac{d}{dr} r \frac{d}{dr}(D_1 n_e) = -w n_e \qquad \ldots\ldots (11\,a)$$

and

$$\frac{1}{r}\frac{d}{dr} r \frac{d}{dr}(D_2 \ln n_i) = e n_i (1-\gamma^{-1})\epsilon_0^{-1} \qquad \ldots\ldots (11\,b)$$

where

$$D_1 = \gamma(D_i b_e + D_e b_i)/(b_e + \gamma b_i) \qquad \ldots\ldots (11\,c)$$

$$D_2 = (D_i\gamma - D_e)/(b_e + \gamma b_i). \qquad \ldots\ldots (11\,d)$$

Equation (11 *a*) is the conservation equation, and is thus the suitable starting point to find the radial variation $n_e(r)$. Equation (11 *b*) is Poisson's equation, and hence suitable for finding $n_i - n_e$ or γ. Naturally, with the approximations made, (11 *a*) and (11 *b*) cannot both be satisfied exactly. After we have obtained $n_e(r)$ from (11 *a*) we must therefore then adjust the value of γ so as to give the best approximation to the true discharge parameters. To do this we calculate γ by a method which is equivalent to that of least squares, from the requirement that the integral of the differences of the right-hand and left-hand sides of (11 *b*) over the cross section of the tube vanishes. To solve these equations we must from now on assume that b, D and w are determined primarily by E_z and not by E_r, so that average values independent of r may be used for these quantities.

The simplest solution of (11 *a*) satisfying the boundary conditions (8 *c*) is

$$n_e = n_{e,0} \mathrm{J}_0\{(w/D_1)^{1/2} r\}. \qquad \ldots\ldots (12)$$

To satisfy (8 *d*) w must be given by

$$w = D_1 (A/R)^2 \qquad \ldots\ldots (13)$$

where A is the first zero of J_0. Solutions involving higher zeros of J_0 are possible, but will be ignored, as in the normal Schottky theory.

When $\gamma = 1$, $D_1 = D_{am}$ and (13) becomes

$$w = D_{am}(A/R)^2 \qquad \ldots\ldots (14)$$

where D_{am} is Schottky's normal ambipolar diffusion coefficient. In this case E_z is independent of j.

Integrating (11 *b*), and introducing the expression for I from (9), with allowance for the boundary conditions (8 *c*), gives

$$\left(r\frac{d\ln n_i}{dr}\right)_{r=R} = \frac{(\gamma-1)I}{2\pi\epsilon_0 E_z (D_i\gamma - D_e)}. \qquad \ldots\ldots (15)$$

Here the apparent difficulty arises that the expression would be infinite if the boundary condition (8 *d*) were rigorously true. This is however not the

case, as (8*d*) gives only a convenient approximation sufficient for the previous calculations. In (15), however, we must use the more exact boundary condition

$$(n_i v_i/4)_{r=R} = -(D_i\, dn_i/dr)_{r=R} \qquad \ldots\ldots(16)$$

where v_i is the mean speed of the ions. Inserting (16) in (15) gives

$$\gamma = \left(1+\frac{q}{I}\right)\Big/\left(1+\frac{D_i q}{D_e I}\right) \qquad \ldots\ldots(17\,a)$$

where

$$q = 2\pi\epsilon_0 R E_z v_i D_e/(4D_i). \qquad \ldots\ldots(17\,b)$$

Equation (17) gives immediately the departure from quasineutrality as a function of field and current. It provides therefore a means of defining formally the boundary between the regimes where the assumption of quasineutrality is valid, and where it fails. We expect that this boundary, which is not sharp, will separate the normal and subnormal regimes. If, for example, a subnormal column is defined to be one for which there is more than a 25% departure from quasineutrality, the largest current I_g for which the column is subnormal will be given by

$$q/I_g = 1/4. \qquad \ldots\ldots(18)$$

In addition (17) leads to the interesting result that γ approaches the limit D_e/D_i as the current density is diminished, so that D_1 and D_e then become identical, that is, for very small currents—actually lower than those of the figure—the longitudinal field approaches a limit E_0 obtained by replacing D_{am} by D_e in Schottky's equation (14).

With equations (11*c*), (13) and (17) we have at the same time achieved our aim of finding the general current–voltage characteristic; if the value of γ from (17) is inserted in (11*c*) and (13) we obtain, since w is, at least in principle, a known function of the field strength,

$$w(E_z) = D_1(I)\{A/R\}^2 \qquad \ldots\ldots(19)$$

which gives the characteristic.

§ 4. Comparison with Experiment

To apply the results of § 3 to the experiments of Mierdel and Schmalenberg we need the values of w, b_e, b_i, D_e and D_i, or of the free paths λ_e and λ_i, and the energy distributions, from which the first set of quantities can be calculated. The latter will be taken to be Maxwellian, with temperatures T_e and T_i. Unfortunately, direct measurements of these quantities in the required ranges are not available, and we must obtain them indirectly. From standard data for free paths (D'Ans and Lax 1949), allowing for the Ramsauer effect (von Engel and Steenbeck 1936) and charge exchange (Weizel and Rompe 1949), the values adopted for the free paths for the conditions of Mierdel and Schmalenberg's experiments have been: for λ_e, $9{\cdot}5\times10^{-3}/p$ cm/mm Hg, and for λ_i, $3{\cdot}2\times10^{-3}/p$ cm/mm Hg.

From consideration of the increase of energy of an ion in a free path and the loss of energy in collisions it is clear that T_i will not be significantly different from the gas temperature. T_e will however be considerably greater. It is given in general by*

$$T_e = ge\mathrm{E}\lambda_e/\{2kf^{1/2}\} \qquad \ldots\ldots(20)$$

* This formula and those for D, b, etc., used for later calculations are all elementary or straightforward adaptations of standard relations, given for example by Loeb (1939).

where f is the average fractional energy loss of an electron in a collision with an atom, which is itself a function of T_e, and g a number of order unity. Under the conditions of Mierdel and Schmalenberg's experiments f will be determined largely by inelastic collisions. In the range of values of present interest f definitely does not have the low value $f_e \simeq 2m/M$ for elastic collisions. This is shown by a measurement at a field of 0·17 v cm^{-1} and pressure of $1{\cdot}9 \times 10^{-2}$ mm Hg, which gave an electron temperature of 8500°K (Knoll *et al.* 1936). Calculation of T_e from (20) under these conditions using f_e gives a value of a greater order of magnitude. The calculation of f can be much simplified by using Elenbaas's model of the mercury atom, with a single effective excitation potential V_a, which will be taken to be 7·8 v (Elenbaas 1935, 1951). More detailed calculations show that the uncertainty introduced by use of this model is without significant effect on the main results. f can then be found like an ionization coefficient by integration of the product of the energy distribution and excitation functions. As the mean electron energy is small compared with eV_a this gives

$$f = A \exp(-eV_a/kT_e) \qquad \ldots\ldots(21)$$

where A is a constant depending on pressure, whose value could be determined from the measurement of electron temperature mentioned above. However, for the following calculations the absolute values of A and g are not needed.

The ionization rate will be given similarly by

$$w = C \exp(-eV_i/kT_e) \qquad \ldots\ldots(22)$$

where V_i is the ionization potential, and C a second constant, completing the quantities whose values are required.

Considering first the limiting conditions for the subnormal discharge, defined by (18) and (17*b*), we find for the limiting current density j_g under Mierdel and Schmalenberg's conditions

$$j_g = I_g/(\pi R^2) \simeq 13{\cdot}3 E_s \,\mu\text{A cm}^{-2} \qquad \ldots\ldots(23)$$

where E_s is the constant field strength in v cm^{-1} for the normal Schottky column at higher currents. The values of j_g are shown in the figure.

To determine the current–voltage characteristics we divide (13), with the value of w inserted from (22) and (21), by the corresponding relation for the normal Schottky column, and obtain

$$(E_z/E_s)^{2V_i/V_a} = \gamma\{1 + (b_i/b_e)\}\{1 + (\gamma b_i/b_e)\}^{-1} \qquad \ldots\ldots(24)$$

using mean values of b and D. Since $\gamma b_i/b_e \ll 1$ for the whole range of conditions studied by Mierdel and Schmalenberg, (24) reduces to the simple result

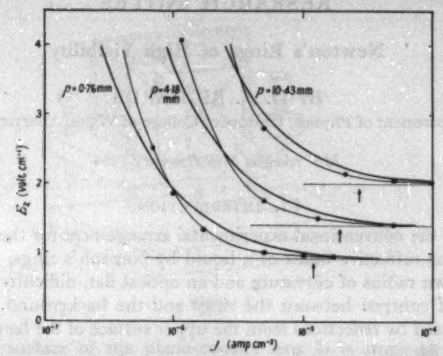
$$(E_z/E_s)^{2V_i/V_a} = \gamma. \qquad \ldots\ldots(25)$$

If we now solve (17) for I, remembering that $D_i/D_e < b_i/b_e \ll 1$, and introducing the expression for γ from (25), we obtain the general equation of the characteristic. Using numerical data found as outlined above, we obtain finally

$$j = 3{\cdot}3_2 E_z/\{(E_z/E_s)^{2{\cdot}66} - 1\}\,\mu\text{A cm}^{-2} \qquad \ldots\ldots(26)$$

where E_z and E_s are measured in v cm^{-1}. The characteristic curves calculated

from (26) are shown in the figure. Considering the uncertainty in the experimental data used in obtaining the numerical constants in (26), the agreement is satisfactory.



Subnormal characteristic curves for mercury at three pressures. The thick curved lines are the computed characteristics, the thin curved lines the experimental characteristics given by Mierdel and Schmalenberg from the measured points shown. The horizontal thin lines are the normal Schottky characteristics. For currents smaller than indicated by the arrows, theory predicts more than 25% departure from quasineutrality j_g.

Acknowledgments

The author wishes to thank Professor Emeléus for suggesting this problem and for his interest in the work during its progress. He is also indebted to the Warren Committee of the Royal Society and the Queen's University of Belfast for making his stay in Belfast possible.

References

D'Ans, T., and Lax, E., 1949, *Taschenbuch für Chemiker und Physiker* (Berlin : Springer)

Elenbaas, W., 1935, *Physica*, **2**, 169, 757; 1951, *The High Pressure Mercury Vapour Discharge* (Amsterdam : North Holland Publishing Co.).

von Engel, A., and Steenbeck, M., 1936, *Wiss. Veröff. Siemens*, **15**, 54.

Knoll, M., Ollendorff, F., and Rompe, R., 1936, *Gasentladungstabellen* (Berlin : Springer).

Loeb, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (London : Chapman and Hall),

Mierdel, G., and Schmalenberg, W., 1936, *Wiss. Veröff. Siemens*, **15**, 60.

Schottky, W., 1924, *Phys. Z.*, **25**, 635.

Weizel, W., and Rompe, R., 1949, *Theorie Elektrischer Lichtbogen und Funken* (Leipzig : Barth).

RESEARCH NOTES

Newton's Rings of High Visibility

By D. A. RICHARDS
Department of Physics, University College of Wales, Aberystwyth

MS. received 11*th February* 1954

§ 1. Introduction

With the conventional experimental arrangement for the determination of the refractive index of a liquid by Newton's rings, using a lens of known radius of curvature and an optical flat, difficulty is experienced due to lack of contrast between the rings and the background. This lack of contrast is caused by reflections from the upper surface of the lens and the lower surface of the flat.

That these unwanted reflections tend to swamp the interference pattern is seen from the following considerations. If I_0 is the intensity of the beam incident on the upper surface of the lens, n_1, the refractive index of the glass of the lens and the flat (assumed equal), and n_2 the refractive index of the liquid trapped between them, then, due to the undesired reflections, a constant intensity of $2[(n_1-1)/(n_1+1)]^2 I_0$ is superposed on the variable intensity due to the interference pattern; the latter varies between 0 and $4[(n_1-n_2)/(n_1+n_2)]^2 I_0$ for normal incidence. In the case of water $n_2 = 4/3$, and if $n_1 = 3/2$ the intensity of the observed pattern ranges from 8% to 9·4% of that of the incident light. These calculations assume that the percentage of light reflected at each face is small and that only one reflection contributes tc the interference pattern.

A method of eliminating the undesired reflections is described by Wood (1924) in which a narrow angle wedge of glass rests on the convex surface of the lens, whose lower surface is coated with a mixture of lamp black and glycerine to absorb reflections from the plane face of the lens. The image of the source of light reflected by the upper face of the wedge does not then enter the microscope. An alternative method in which both undesired reflections are absorbed is described below. This method and the form of the apparatus has been in use in the teaching laboratory of this College for a number of years.

§ 2. Apparatus

The apparatus is shown in the exploded diagram figure 1. ABC and ABD are two 45° prisms; the hypotenuse face of one is partially silvered to reflect about half the incident light downwards, and the prisms are cemented together with Canada balsam to form a cubical block of 1¼ in. edge. The face BD is coated with a mixture of lamp black and Canada balsam to absorb light transmitted through the hypotenuse face. This light would otherwise be reflected back into the microscope objective M.

The complete block is strapped as shown to a vertical plate P attached to a bracket carrying three levelling screws. The feet of these screws rest on a

hole slot and plate arrangement on the base-plate E. In figure 1 the feet of the levelling screws are shown well above E to show the lens assembly more clearly.

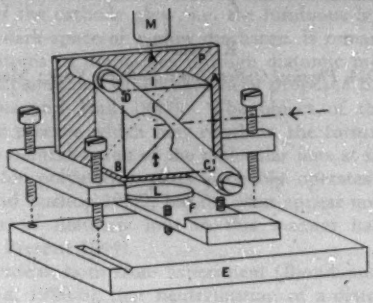


Figure 1.

The plane surface of the plano-convex lens L is stuck to a short brass beam F with the lamp black–Canada balsam cement to prevent reflection from the lower plane surface. The beam F is provided with two screws, the points of which serve as pivots for the beam. The beam is counterpoised so that, in the working position, the curved upper face of the lens presses lightly against the lower face of the glass block. The centre of the ring system is brought to the centre of the lens by adjusting the three levelling screws, which are then locked in position by nuts which are not shown in the diagram. Once this adjustment had been made no further adjustment was found necessary over a long period of time despite the fact that the lens and prism block were removed many times for cleaning the surfaces.

The radii of the rings are measured first with an air film to determine the radius of curvature of the lens surface and then with a liquid film. To obtain the liquid film, the free end of the beam is lifted and the beam withdrawn. A small drop of the liquid is placed on the centre of the lens and the beam is replaced. The free end of the beam is lowered when the centre of the lens is below the centre of the block.

With the arrangement described above, the refractive index of the prisms and lens (of radius of curvature 104 cm) being 1·52, rings have been measured up to a diameter of 2·5 cm—the diameter of the lens—when using a liquid of refractive index 1·48. With the conventional arrangement of Newton's rings the intensity of the ring pattern would fluctuate between 8·51% and 8·58% of that of the incident light. The visibility of the fringes would be too low to be observed and measured with confidence.

Figure 2 (Plate) is a composite photograph of interference rings taken with air and water films with the conventional Newton's ring apparatus (I and III respectively) and with the apparatus described above (II and IV respectively). The light source was the green mercury line. Panchromatic plates were used, the exposures being such as to give approximately equal blackening in the densest parts of each negative. The negatives were printed on the same hard grade of paper.

It will be noted that only in figure 2 (IV) is there a dark spot at the centre, showing that the contact between the two surfaces was imperfect in the other cases, but this is immaterial to the present work, which was to show a method of increased contrast.

Reference

Wood, R. W., 1924, *Physical Optics*, 2nd Edn (New York: Macmillan), p. 154.

LETTERS TO THE EDITOR

Dissociative Recombination at Surfaces

The spectrum of the cathode glow, i.e. the luminous layer at the negative end of the cathode dark space of a glow discharge, is remarkably rich in lines of the neutral gas atoms in discharges through diatomic molecular gases such as nitrogen. The explanations which have been proposed for this are not fully satisfactory (Emeléus and Hall 1931). The object of this note is to call attention to another process which may result in the formation of the atoms, namely, dissociative neutralization of the molecular ions at the cathode surface. A similar process for polyatomic gases probably operates in Geiger–Müller counters (Massey and Burhop 1952), but for what appear now to be inadequate reasons, dissociation of diatomic ions in this manner has been considered unlikely (Korff and Present 1944).

It is however evident, both from experiment (Biondi and Brown 1949) and theory (Bates 1950 a, 1950 b), that neutralization of a diatomic molecular ion by an electron in the gas phase is very frequently followed by spontaneous dissociation. It seems therefore probable that the corresponding process can take place at the cathode surface. It would be represented by

$$AB^+ + e(\text{Metal}) \rightarrow A + B$$

where AB^+ is the incident ion and A and B the corresponding neutral atoms. An electron from a conduction level in the cathode metal passes through the surface potential barrier to a vacant level of AB. If the latter is repulsive, or reached above its dissociation asymptote, dissociation can occur, at least if the molecule is suitably orientated relative to the surface.

The average distance from a metal surface at which an ion is neutralized is not very sensitive to the speed of the ion (Cobas and Lamb 1944). Hence, if this process occurs for positive rays at the cathode of a glow discharge, it can presumably also occur when slow diatomic molecular ions approach any clean metal with suitable Fermi levels and work function. It could, for example, be the origin of the chemical activity suspected to occur in nitrogen which has been activated by electrons with energy slightly above the first molecular ionization potential (Massey and Burhop 1952). A thin non-metallic film on a metal surface, whilst not preventing neutralization, might however modify the processes in such a way as to favour formation of stable neutral molecules, and it is not at all clear what will happen at the surface of an insulator receiving an ambipolar current of molecular ions and electrons from a discharge plasma.

I am indebted to Professor D. R. Bates and Mr. D. T. Stewart for discussion of this work.

K. G. Emeléus.

Physics Department,
Queen's University,
Belfast.
22nd March 1954.

Bates, D. R., 1950 a, *Phys. Rev.*, **77**, 718; 1950 b, *Ibid.*, **78**, 492.
Biondi, M. A., and Brown, S. C., 1949, *Phys. Rev.*, **76**, 1697.
Cobas, A., and Lamb, W. E., 1944, *Phys. Rev.*, **65**, 327.
Emeléus, K. G., and Hall, O., 1931, *Proc. R. Irish Acad.*, **40**, 1.
Korff, S. A., and Present, R. D., 1944, *Phys. Rev.*, **65**, 274.
Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford : University Press).

Cascade Excitation of Phosphors by Ultra-Violet Radiation

An unusual type of photoluminescence has been found in a number of borates and aluminates activated by lead or thallium. A typical case is $5CaO \cdot 3Al_2O_3$ with 1% Tl added (as thallous nitrate) before firing at 1250°c.

Two sources of ultra-violet radiation were used, a ' germicidal ' low pressure mercury discharge lamp with Corning 9863 filter, emitting mainly at 2537Å; and a high pressure quartz mercury arc in a Wood's glass envelope, emitting mainly at 3650Å. When exposed to either source, at about the intensities used for normal phosphor testing, the aluminate gave a feeble yellow luminescence, increasing gradually with time of irradiation. After the 2537Å excitation a slight afterglow was visible, and a marked thermoluminescence at (roughly) 100°c. Exposure to 2537Å, even for a few seconds, followed by a further exposure to 3650Å, produced an immediate bright yellow fluorescence, increasing over a certain range with the length of the first exposure, showing a very slow decrease with time of exposure to the source, with a slight afterglow at room temperature, and having a thermoluminescence similar to that described above except that it decreased with length of exposure to 3650Å. Exposure to the two sources in reverse order gave no similar effects. The dark period between the exposures could be extended to at least a week with only partial loss of the output under 3650Å, though this loss may be partly due to deterioration of the hygroscopic phosphor, a property which seemed to affect a number of the experiments described.

In view of the apparent similarity of the effect to ' radio-photoluminescence ' (Furst and Kallmann 1951), the effect of an initial brief exposure to 45 kv x-rays was tried, and found to render the powder sensitive to subsequent 3650Å radiation, but not to 2537Å radiation.

Under more intense sources, used to enable spectrophotometric estimation of the output to be made near the peak of the emission band, the powder being in a metal cell with a quartz window, either source alone gave a rise of luminescence to a maximum over about 20 minutes, whereas together the (increased) maximum was reached in about 5 minutes. In this case the 2537Å source was a quartz mercury arc without filter. The emission declined in each case at a very slow rate during continued excitation, and also in the dark, for recommencement after a resting period gave values of the light output falling roughly on the lines indicated in the exposed periods. When the initial experiment of successive exposures to 2537Å and 3650Å was tried in the spectrophotometer, a light output pattern strongly reminiscent of infra-red stimulation was produced. After exposure to 2537Å (2 or 75 minutes in separate trials), the 3650Å lamp produced an immediate high emission, fading within a few minutes to the value for the normal curve for 3650Å alone. This rapid rise and fall was not seen in the samples inspected by eye under the lamps. The light sources were not by any means monochromatic, and some of the effects are probably complicated by this as well as by the deterioration due to moisture.

The spectral regions responsible for the two stages of excitation were determined by a further experiment. A screen of the phosphor exposed in the focal plane of a Hilger medium quartz spectrograph to an iron arc spectrum for 30 minutes gave no visible emission, but on shutting off the spectrum and exposing the whole plate to the 3650Å lamp the iron spectrum was at once revealed in

yellow luminescence from about 2300 to 3000 Å. The converse experiment, with the whole screen first exposed to the 2537 Å lamp and then to the iron arc spectrum, showed the excitation region for the second stage to be a band extending from about 3000 to 4000 Å. However, on continued exposure to the spectrum the lines between 2300 and 3000 Å also appeared slowly, in marked contrast to the immediate appearance of the longer wavelengths.

The emission spectrum of the calcium aluminate was a band with its peak at 5900 Å and nearly symmetrical on a frequency scale, with a width at half peak intensity of 1400 Å. Two small peaks at about 6700 Å and 6400 Å were superimposed on the main band.

The powder exposed to 2537 Å gave the thermoluminescence described above, but without loss of the power to respond to 3650 Å after cooling. To remove all evidence of excitation by 2537 Å from the phosphor it was necessary to heat it to a high temperature. Marked loss of sensitivity to the cascade effect occurred after heating to 500°c, but about one hour's heating at 580°c was necessary to remove completely the latent effects of the exposure to 2537 Å. The deleterious effects of moisture could also be removed by heating.

The behaviour of the phosphor has resemblances to radio-photoluminescence, though the quanta used in the first stage of excitation are of a much lower order and scarcely seem sufficient to produce new emission centres. It also resembles stimulation by infra-red in some respects. It is clear at any rate that the phosphor contains both shallow and very deep metastable levels. Crucial experiments to elucidate the mechanism will require more nearly monochromatic irradiation of samples sealed in evacuated quartz tubes, and experiments on these lines are in progress. It will be important to determine the effects of different intensities of the two sources, and also of their relative intensities, since the results so far obtained seem to be complicated by differences in this respect.

Thorn Electrical Industries Ltd.,
Great Cambridge Road,
Enfield.
25th March 1954.

P. W. Ranby.
S. T. Henderson.

Furst, M., and Kallmann, H., 1951, *Phys. Rev.*, **82**, 964, **83**, 674.

The Cause of Stress in Evaporated Metal Films*

Metal films condensed from the vapour in high vacuum are observed to be in a state of relatively high tensile stress (Crittenden and Hoffman 1950). Recently Murbach and Wilman (1953) have measured stresses for several metal films and have proposed a mechanism for the origin of the stress. Their interpretation is that the films are relaxed to a stress-free condition at roughly the recrystallization temperature during deposition. Subsequent cooling to substrate temperature then gives rise to the stress through thermal expansion. The purpose of this letter is to point out an alternative mechanism which has been developed from experiments in this laboratory.

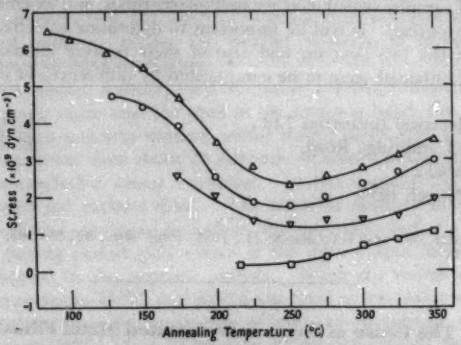
* Work supported by the U.S. Office of Naval Research.

Evaporations have been carried out for Ni films deposited on substrates held at temperatures of 75°, 125°, 175° and 225°C. The substrates were clamped against a lead backing to insure good thermal contact. The films were about 1500 Å thick and were formed at a rate of 60 Å sec^{-1} in a residual pressure of 1×10^{-6} mm Hg. This rate and pressure give a maximum of 1% contamination if all the residual gas atoms that hit the film surface are trapped. A good vacuum is necessary since the change in stress on annealing is particularly sensitive to the presence of water vapour during the time of evaporation.

The stress in such films is measured by observing the curvature of the thin mica substrate supported on knife edges at the ends. The stiffness constant for the composite beam of substrate and film is determined by loading the samples at the centre and measuring the resulting change of curvature.

The measurements of stress were made with the substrate at 25°C. The curvature of the substrate at this temperature can be described as due to two stresses. One is the result of differential thermal expansion of the film and substrate in cooling from the substrate deposit temperature to 25°C. The second, which we shall call the intrinsic stress, results from irreversible internal changes in the film. The differential expansion component is evaluated by measuring the curvature at the deposition temperature and again at 25°C. For Ni evaporated on mica held at 75°C it amounts to $0{\cdot}4 \times 10^9$ dyn cm^{-2}, in good agreement with that calculated from the expansion coefficients.

The intrinsic stress has been measured as a function of both substrate deposit temperature and annealing temperature. The figure shows the intrinsic



Variation of the intrinsic stress in evaporated nickel films as a function of annealing temperature. The temperature of the initial point of each curve represents the substrate temperature during deposition. All values measured at 25°C

stress observed at 25°C for four different films after 15-minute anneals at the temperatures indicated. The lowest temperature point for each curve corresponds to the deposition temperature for that sample. It is seen that, on annealing, the intrinsic stress irreversibly decreases and reaches a minimum value at about 250°C, and then rises. This behaviour is inconsistent with Murbach and Wilman's model, since relaxation of the film at roughly the recrystallization temperature (for Ni, 600°C) should make the film stable under

annealing at lower temperatures. Observation of the electrical resistivity also indicates irreversible changes whenever the temperature first exceeds the substrate temperature maintained during deposition (Hoffman, Anders and Crittenden 1953). The physical mechanism whereby a surface layer could be maintained at a temperature such as 600°c while the substrate is at a much lower temperature, say 25°c, seems difficult to conceive. The incident radiant energy from the vapour source plus the energy received by the surface by virtue of the arriving atoms would give a temperature drop of about 10^{-5} deg. c through a 1000Å film, taking the thermal conductivity value for bulk nickel. Arrival of an individual atom at the surface will release an amount of energy given by the kinetic energy of arrival plus the binding energy. Normally the arriving atom will migrate a short distance on the surface before settling at an attachment site. Neglecting this, and examining the hypothetical situation at an instant at which the energy is shared by the first and second neighbours of a site, one finds a 'temperature' rise of roughly 700°c. This idealized 'temperature' pulse would disappear by thermal conductivity in hemispherical geometry with a relaxation time of the order of 10^{-12} second and would hardly produce stress relaxation by conventional recrystallization.

An alternative mechanism for the origin of the intrinsic stress, which seems to fit the physical phenomena better, is that imperfections formed during deposition undergo change shortly after being trapped by succeeding layers of deposited metal. These changes take place at essentially the substrate temperature, perhaps with some help from the temperature pulses due to arriving atoms. Sufficient information to identify the imperfections with certainty is lacking at present. However, it seems likely that the irregular regions on the growing surface, which become crystal boundaries when buried, rearrange and give rise to the principal part of the intrinsic stress through a reduction of volume. Imperfections may also be trapped at points in otherwise relatively perfect crystals, and changes in some of these may also contribute to the intrinsic stress. The variation of the initial intrinsic stress with substrate temperature during deposit, as shown in the figure, is consistent with either of these models. Increasing the substrate temperature should increase the surface mobility of arriving atoms and decrease the probability of attachment at an abnormal surface site. This should lead to a lower initial imperfection density and would explain the lowered intrinsic stress.

Average values of stress in a few other metals at several deposition temperatures are quoted in the table. It is interesting to note that for Cu the intrinsic stress is very low. In Cu the annealing temperatures for bulk metal are low compared with Ni and the surface mobility is known to be high compared with Ni. Observations of the behaviour of Cu in agglomerating when deposited in small thickness gives the evidence for surface mobility.

Stresses in Several Metals as a Function of Deposit Temperature

Metal	Ni	Ni	Cu	Cu	Fe*	Co*
Substrate deposit temp. (°c)	75	125	25	75	90	200
Total stress (10^9 dyn cm^{-2})	6·80	4·70	0·18	0·37	5·5	4·9
Diff. exp. stress (10^9 dyn cm^{-2})	0·40	0·94	0·0	0·33	0·5	1·5
Intrinsic stress (10^9 dyn cm^{-2})	6·40	3·76	0·18	0·04	5·0	3·4

* Water vapour present during evaporation.

At very low deposition temperature the changes responsible for volume reduction may not be thermally activated during deposition. In this case the intrinsic stress should be low and should increase on annealing at low temperature. Such effects are being investigated.

A mechanism for the annealing behaviour after deposition at intermediate temperatures has been previously proposed (Hoffman, Anders and Crittenden 1953). This is based on the idealized decay of imperfections within the crystals. This mechanism considers the films in the deposited state to contain a large number of lattice vacancies. These imperfections migrate in the metal with a rate dependent on the temperature of the sample. When a vacancy is filled, by virtue of diffusion of the vacancy to coalesce with other vacancies, the neighbouring atoms of the former vacancy site relax outward. This causes the crystals to expand, reducing the tensile stress. As the vacancies coalesce into larger aggregates, they finally collapse to form dislocation rings. This causes contraction of the crystals and an increase of stress. This sequence is consistent with the shape of the curves in the figure.

Earlier work on annealing, after deposition at room temperature and above, showed only an increase in stress (Crittenden and Hoffman 1950). This was related to gas trapped in the film during evaporation in relatively poor vacuum (10^{-4} mm Hg). Under such conditions part of the initial intrinsic stress may also result from effects of gas trapping. Water vapour has been found to produce this effect more strongly than N_2 or O_2. If deposition is carried out at pressures of 10^{-6} mm Hg or better, at rates of 60 Å sec^{-1} or more, the effect is absent.

The intrinsic stress present in a film under given conditions is thus believed to be the result of the history of its imperfections. This includes the crystal boundaries, as well as imperfections within the crystals that make up the polycrystalline film. The total stress usually includes a component resulting from differential expansion relative to the substrate, normally small compared with the intrinsic stress. Full understanding of the properties of the imperfections awaits the existence of much more complete information on a variety of metals. Information drawn from irradiation effects and from low temperature plastic deformation of metals promises to throw light on the problem.

Case Institute of Technology,
Cleveland,
Ohio, U.S.A.
15th March 1954.

R. W. Hoffman.
R. D. Daniels.
E. C. Crittenden, Jr.*.

Crittenden, E. C., Jr., and Hoffman, R. W., 1950, *Phys. Rev.*, **78**, 349.
Hoffman, R. W., Anders, F. J., and Crittenden, E. C., Jr., 1953, *J. Appl. Phys.*, **24**, 231.
Murbach, H. P., and Wilman, H., 1953, *Proc. Phys. Soc.* B, **66**, 911.

* Now at the U.S. Naval Postgraduate School, Monterey, California.

REVIEWS OF BOOKS

Probability and Information Theory, with Applications to Radar, by P. M. WOODWARD. Pp. x+128. (London: Pergamon Press, 1953.) 21*s*.

One of the merits of modern statistical communication theory springs from its breaking down of the last vestiges of empiricism in telecommunication engineering. Prior to its development, the various operations of *detection*, *filtering*, *coding*, *noise-reduction*, etc. were based upon arbitrary parameters of design—such as *bandwidth*, *signal-to-noise ratio*, etc.; nevertheless the principles of design (network theory), which still employ these parameters in practical cases, are mostly of great elegance. But the empiricism was there—and, to many people, irksome.

The revolution that was brought to our subject, largely through the vision of Wiener and the mathematical industry of Shannon, came through the introduction of probability theory. Philosophically, this is directly in line with the lesson of Locke and Hume: that all 'knowledge' of the physical world we gain through observation is probabilistic. And anyone who is engaged in, say, conversation or otherwise communicating with another, is certainly observing a phenomenon of nature (albeit a unique one). The whole study of communication has become more rational and intelligible through the introduction of probability theory.

The author of this book has wisely started by sketching the outlines of (statistical) probability theory, at a level suited to the beginner, dealing with the binomial, the Poisson and the gaussian distributions. For the beginner, a few more examples of typical physical situations corresponding to such distributions would have been helpful. Further introductory material deals with different aspects of signal analysis; for example, the use of moments and generating functions, the processes of convolution and of sampling (reduction of band-limited wave-forms to finite sets of discrete data, and use of interpolation functions).

The central sections of the book survey the principal contributions of Shannon to statistical communication theory, especially the definition of *information* as *statistical rarity*, and his channel capacity formula; incidentally, it is a pity the author uses the expression 'information theory' which is more usually associated, in this country, with the logic of experimentation. Finally, interpretation is given to the communication process, when noise is present, in terms of the theory of inverse probability—briefly, a receiver can at best present only a likelihood function, on the receipt of physical and noisy signals, giving relative weightings to the probabilities that various alternative messages may have been sent. The final sections illustrate and expand this whole approach by reference to various problems of radar, especially the problem of detection of pulses in noise.

The book may be thoroughly recommended; we have, as yet, few books on this subject in Britain. It should be comprehensible and of interest to anyone familiar with the classical Fourier approach to telecommunication. The only criticism which the reviewer would make concerns certain details of the mathematical formulation and a general regret that the illustration of radar puts a restriction on interpretation which the very general treatment of the book does not deserve.

COLIN CHERRY.

Microwave Lenses, by J. BROWN. (Methuen's Monographs on Physical Subjects.) Pp. xi+125. (London: Methuen, 1953.) 8*s*. 6*d*.

This monograph will be welcomed both by workers in the microwave field and by those interested in the relatively new and important field of microwave optics. As is the intention of these monographs, the book presents a clear and concise account of the present position as regards microwave lens design, and the author is to be commended on the arrangement and presentation of the subject.

The first two chapters deal with aerial design by optical methods and the application of microwave lenses, together with the requisite treatment of diffraction, amplitude distributions, and phase errors. These are followed by chapters on solid dielectric lenses, metal delay dielectrics, and metal plate and rodded dielectrics. Amongst the topics of fundamental importance dealt with here are the impedance concept, surface reflections, the application of the Debye dielectric theory to delay dielectrics, the transmission line theory for delay dielectrics, and the reflections from metal plate media. These complicated topics are dealt with in a remarkably clear and concise manner. Practical questions such as tolerances and bandwidth in the various media also receive due consideration, as well as the various methods of lens construction.

Next follows a discussion of the beam distortions introduced by moving the primary feed from the lens focus, as in the problem of wide angle scanning. Expressions are given for the phase errors introduced, together with the maximum scanning angle for a satisfactory radiation pattern, and the locus of the primary feed. The advantages of using non-homogeneous lenses in which the refractive index varies continuously with position are also dealt with in connection with the possibility of obtaining a 360 degree scan by moving only the primary feed. Expressions are given for the required variation of refractive index in such systems as the Maxwell fish-eye and the Luneberg lens. Practical methods for the construction of such media are given, and as is general throughout the book, the theory is well supplemented with experimental data.

The remaining type of lens dealt with are path length lenses in which the physical length of each path from a point feed to the aperture plane is kept constant, whence no refracting medium is required. Finally the book closes with a chapter on the design of lens aerials in general, and the important topics of when to use a lens aerial, the type to be preferred for a given application, and the production of beam shapes other than so called ' pencil beams '.

This wide field of microwave lens design is covered extremely well for a book of this size and is supplemented by an excellent bibliography. The book is very well illustrated with 53 figures and 4 plates, and the printing and presentation conform to the usual high standard of these monographs. W. CULSHAW.

Properties of Metallic Surfaces. Institute of Metals Monograph and Report Series, No. 13. Pp. iv+368. (London: Institute of Metals, 1953.) 35*s*.

In many studies of the solid state it suffices to characterize our working materials by their bulk properties—density, strength, electrical and thermal conductivity, etc., and to describe their behaviour entirely in terms of these properties. In recent years, however, it has come to be more widely realized that in many practical problems a precise description of the behaviour of a finite

volume of a solid can be obtained only by taking into account effects arising at the discontinuity formed by its free surface. The study of solid surfaces has two aspects. Firstly, in deriving our information about the bulk properties, we often employ information gleaned from a study of the surface, as for example in studies of plastic deformation. It is important therefore to find out how truly the two-dimensional evidence on the surface reflects the three-dimensional properties of the solid and also to what extent the surface may itself affect the bulk behaviour of the solid when, for example, the surface to volume ratio of the specimen becomes large. Secondly, there are those properties of a surface which are of a more truly interfacial nature, arising from the interaction between the surface and other phases, solid, liquid or gaseous brought into close proximity with it. The phenomena observed, particularly in the second category, are often complex and seldom amenable to rigorous mathematical treatment and demand for their full understanding a wide knowledge of many disciplines—metallurgy, physics, chemistry, electrochemistry, etc.

The Institute of Metals is therefore to be congratulated on its bold initiative in bringing together for the symposium, of which this monograph is a report, experts from many of the diverse specializations which have contributed to our knowledge of metallic surfaces. If the discussion at the symposium was at times disjointed, this was to be expected when the subject matter ranged from the practical problems of running bearings to the high-frequency resistance of surface layers.

The main papers, numbering thirteen, differ considerably in emphasis. Some, for example, concentrate mainly on reviews of the author's own work in a particular field (Andrade on the effect of surface conditions on the mechanical properties of metals, Bowden and Tabor on the influence of surface films on the friction and deformation of surfaces). Being highly individual, these open the way for a controversial and fruitful discussion. Of the review papers, those aiming to give a critical survey are to be preferred over the expanded card index. The latter, while useful as source material (one just fails to top 200 in the number of references given) is of little value to a worker outside a field who requires a critical succinct evaluation of the current position in that field. A number of papers on individual research projects are included in the monograph. The value of these in a symposium of this nature is doubtful. Although it is an excellent example of a careful piece of work on a difficult problem, the paper on machining and grinding by Spear, Robinson and Wolf hardly merits forty pages here. In view of the wide range of interests of the readers, a summarized account of the main experiment and conclusions would have been adequate, with fuller details reserved for the appropriate journal. It is pleasing to the more academic scientist to find the results of his work making useful contributions to the understanding of industrially important processes, as illustrated by the contributions of Barwell on lubrication and the nature of the superficial layer after prolonged periods of running, Philips on the anodic film on aluminium and its alloys and Evans on chemical behaviour as influenced by surface condition.

The number and variety of the contributions to discussion show that the study of metal surfaces is a very lively field of research. It is to be hoped that the cross-fertilization of ideas arising from this symposium will aid workers in the field towards a fuller appreciation of the value of lines of attack on their problems other than those to which they may have become accustomed in the

past. It is perhaps regrettable for those not present at the discussion that the full fire of some of the verbal exchange did not find its way to the printed page. In this respect the contributor to discussion must regard himself as more fortunate than the contributor to Hansard.

The production of the monograph maintains the customary high standard of the Institute. It can be recommended to all who are interested in solids and the properties of their surfaces. J. W. MENTER.

Introduction to a Study of Mechanical Vibration, by G. W. VAN SANTEN. Pp. xvi + 296. (Eindhoven: Philips Technical Library; London: Cleaver-Hume Press, 1953.) 35*s*.

Nowadays there are a number of first-class books on the subject of mechanical vibrations, and the criteria by which a new text is judged must therefore be stringent. One obvious justification for a new book on any topic is that it should contain material which is not easily available elsewhere and, in this respect, the book under review scores good marks since it gives, in the last four chapters, an authoritative account of the techniques developed by the engineers of the Philips organization at Eindhoven for the study of vibration phenomena. It is true that few references are made to equipment which has been developed outside Holland, but this is not a serious fault, since adequate descriptions are already available.

Since this book is an English version of the Dutch edition which appeared in 1950, it is reasonable to suppose that it should be free from major errors of fact and of principle. This is far from being the case, and the text is marred by a number of glaring fundamental errors. To take a few examples, on page 104 an expression is derived for the velocity of transverse waves in *a rod*, whereas, in point of fact, the treatment and the result apply only to transverse waves in *an infinite medium*. Again, on pages 112 to 113, the treatment of the vibrations of open and stopped organ flue pipes leads to results which are contrary to fact, at least when ordinary nomenclature is used, e.g. " the fundamental tone of an open organ pipe is half the natural frequency of the stopped pipe, and therefore one octave lower ". This and other errors in the same section are the result of regarding the mouth of an organ pipe as a node. In addition there are a number of ambiguities (e.g. in the treatment of reflection and refraction of waves on pages 96–98), and of errors of fact, e.g. on page 136, where it is stated that longitudinal seismic waves travel " at a speed of about 5 miles an hour or 7 feet per second "; observed values vary with geological structure, and they can be as high as 20 000 feet per second.

The translation gives the impression that it is literal rather than idiomatic and there are a number of obvious errors, e.g. ' fixed medium ' for ' solid medium ' on page 95. Finally, the proof-readers or the printers have been negligent; the number of misprints etc. is greater than one normally encounters and, on page 241 in the review copy, the first three lines at the top of the page should be transferred to the bottom.

As it stands this book cannot receive high commendation, neither should it be condemned out of hand; it does show a number of good features and the general plan of the treatment is well balanced. Should there be a call for a new edition it is to be hoped that it will receive the thorough revision which it merits. R. M. D.

Magnetic Amplifiers, by GEORGE M. ETTINGER. Pp. viii + 88. (London: Methuen, 1953.) *6s. 6d.*

This book is one of the well known series of Methuen's Monographs on Physical Subjects which is intended for University students and others. It is therefore of prime importance that a great deal of care and attention should be given to clear expositions of the fundamentals of the subjects considered. Unfortunately in the present monograph the fundamental principles of operation of magnetic amplifiers and related devices are not made clear and, in fact, there are some statements which would lead to considerable perplexity on the part of the uninitiated reader. For example, in the footnote on page 16 there occurs the remarkable statement that " It is well known from basic a.c. theory that the mean voltage across an inductance is proportional to the maximum flux density reached by the magnetic core ". The student who believes that the mean value of an alternating voltage is zero may well be excused if he forms the impression that the operation of magnetic amplifiers is incapable of rational explanation. The later sections of the book are much more satisfactory and they are concerned with brief descriptions of typical practical applications and the limitations of magnetic amplifier devices. These sections are illustrated with well chosen basic circuit diagrams and the graphical analyses of the circuits are completely satisfying if the basic principles are known and understood.

A short table is included as an Appendix which summarizes the relevant properties of some magnetic materials suitable for use in magnetic amplifier systems. The bibliography contains 95 references to articles and books. R. S.

Thermionic Valves, their Theory and Design, by A. H. W. BECK. Pp. xvi + 570. (Cambridge: University Press, 1953.) *60s.*

This book deals with the physics of valves, as distinct from circuit electronics on the one hand and manufacturing technology on the other. It is divided into three sections, the first dealing with electron emission of all kinds and with luminescence. The second describes the interaction between electric and magnetic fields and electrons under all conditions in valves and cathode-ray tubes. The third utilizes the results of the first two sections to describe all kinds of valves and some special purpose tubes.

The subjects are in the main dealt with in two stages, starting at a fairly advanced level, and then proceeding to give an account of the most recent work. The author expresses his own opinions on the applicability of recent publications. The book enables anyone entering a field of research on valves, until he forms opinions of his own, to have the advantage of an expert and balanced point of view to guide him. Ample references to recent work will be found.

A very valuable feature of the book is that it collects together much work on the behaviour of valves at very high frequencies.

Having stated that the book is a very good one we have to mention its shortcomings. These are to be found, not in the subject matter and method of treatment, but in a lack of care and attention to detail. There are some mistakes of fact. For example, on page 225 the Poisson distribution of probabilities is approximated by ' the ' normal distribution. Now the variance of a Poisson distribution is equal to and uniquely determined by the mean value, whereas

for a normal distribution, the variance is not equal to the mean value, as the author states, nor has it any relation to the mean value. The approximation used is therefore wrong. There are also many errors which have escaped proof reading. For example, on page xvi the dielectric constant of free space is given as $36\pi \times 10^{-9}$ farads per metre instead of $10^{-9}/36\pi$ farads per metre. On pages 194 to 196 a succession of errors in regard to suffixes appears. Here the suffix k refers to one of n electrodes. In one equation k appears twice as a multiplier, and in another n appears once as the upper limit of a summation, and also as unit normal vector. Finally, n changes without warning to N. The list of symbols is by no means complete, and is not consistently used.

C. S. BULL.

Foodstuffs, their Plasticity, Fluidity and Consistency, edited by G. W. Scott Blair. Pp. xv+264. (Deformation and Flow: Monographs on the Rheological Behaviour of Natural and Synthetic Products.) (Amsterdam: North-Holland Publishing Company, 1953.) 50*s*.

The object of industrial rheology is to discover what criteria govern the behaviour of materials during manufacture and those which affect the quality of the product as assessed by the consumer. This entails detailed physical measurements which are subsequently replaced by more readily performed tests to control production. Due to the variable nature of the raw materials used, combined with the stringent demands of customers, few industries offer more scope for this type of rheological work than the manufacture of foodstuffs. This book provides the first extensive description of rheology applied to the foodstuffs industry and shows to what degree this application has been successful. The subject matter is divided into eight chapters, each written by different authors, several of whom are research workers in Holland.

Chapter I, on Starch, is written by J. Hofstee and A. H. A. de Willigen of the Experimental Station for Potato Research Processing, Groningen. The authors describe in detail the influence of concentration and the addition of electrolytes on the pasting properties of potato starches. The chemical composition and physical properties of several varieties of starch are contrasted and the chapter concludes with a discussion of the modification of starch by wet methods. The second chapter, entitled ' Cereals ', deals with bread making and dough rheology. Several dough testing instruments are described, with an explanation of how typical records are interpreted. The authors, D. H. Greup and H. M. R. Hintzer, also discuss the rheological behaviour of bread crumb when fresh and during staling. Following a short review of the rheology of milk and allied products by the editor, H. Mulder gives an account of the flow properties of butter and the methods used to investigate them. The relationships between the various methods are tabulated, although it is stated that as yet no rheological measurements give the same result as those of trained judges in butter grading.

A review of the vast amount of work published on the rheology of cheese and curd is given by Miss M. Baron and Dr. G. W. Scott Blair. An example of rheology applied to a naturally produced foodstuff is contained in an excellent chapter on the rheology of honey by J. Pryce-Jones. Honeys are grouped into Newtonian, thixotropic and dilatant types and their properties are clearly

illustrated by appropriate figures and tables. The flow properties of a large variety of food products, e.g. boiled sweets, jellies, marshmallows and chocolate, are logically dealt with by H. G. Harvey, who divides them into single-phase systems, gels, two-phase and multi-phase systems. Finally a chapter entitled ' Psycho-Rheology ' outlines attempts to analyse human assessment of foodstuffs in terms of objective measurements. The very large number of footnotes makes the reading of this chapter unnecessarily tedious.

This book is written in a clear and concise manner and has few misprints. It should prove of interest to rheologists and its extensive bibliography makes it extremely valuable to both students and research workers. J. J. BENBOW.

Essentials of Microwaves, by R. B. MUCHMORE. Pp. vi+236. (New York: John Wiley; London: Chapman and Hall Ltd., 1953.) 36*s*.

The object of the author of this book has been to provide, as simply as possible, an account of the underlying physical principles of microwave devices on the understanding that such principles can be grasped fairly easily by the young student. The presentation of the subject with adequate mathematical background has thus been avoided and in a short book of 236 pages there are over two hundred diagrams and illustrations.

The author, in his preface, gives no information for which class of reader his book is intended, and the present writer has some difficulty in assessing its value to the English student. While the presentation is clear, it is certainly not deep enough for the honours student of physics and engineering who may be specializing in the subject during his final year. As a brief introduction for the first or second year student Dr. Muchmore's book may serve a more valuable purpose, and could with advantage be added to the list of books for recommended reading—if indeed it is ever possible to persuade students to read outside the prescribed limits of degree examinations.

The contents of the book may conveniently be summarized by the following list of chapter headings: Maxwell's Laws; characteristic waves and wave guides; cavity resonators and filters; characteristic waves and antennae; typical microwave antennae; waves and electron streams; grid control tubes; klystrons; travelling-wave and multiple stream tubes; magnetrons; electrical noise; microwave radio systems-relays; radar; applications of microwaves in physical research; microwave measurements. There is no name index.

J. T. R.

The Molecular Architecture of Plant Cell Walls, by R. D. PRESTON. Pp. xii+211. (London: Chapman and Hall Ltd., 1952.) 36*s*.

In recent years the term biophysics has been used rather loosely and sometimes seems to connote almost anything a physicist may choose to do if he enters the field of biology. Professor Preston in this volume has, however, covered in a most attractive way a restricted field of orthodox biophysical research in which he is an outstanding expert. He has given us an account of research, to which he has contributed very largely, into the studies of the cellulose wall of plant cells. The book has been written chiefly for botanists; and in consequence contains a number of chapters outlining not only the chemical nature of the secondary wall, but the physical methods—chiefly those of X-ray diffraction and polarized light—employed in such studies.

To the physicist with any interest in biology the story unfolded in the later chapters is of great interest. To the physicist with no such special interest the book is of value in bringing home to him the revolution which is taking place at the present time in biological thought. This revolution has of course been going on for a long time, and is due to the impingement of other disciplines, particularly chemistry, biochemistry and physics, on the techniques and thoughts of biologists. Professor Preston is one of the growing number of biophysicists who believe that until the structure of living material is fully understood there can be no proper understanding of function. An attractive feature of the book is that while modern physical methods are fully exploited, the work of early pioneers, such as Malpighi and Grew and, much later, von Mohl and Nägeli, has not been forgotten.

To the more physically minded of his readers Professor Preston has not in Chapter II brought out sufficiently clearly the differences on the one hand between plant and animal cells; nor, on the other hand, the distinction between the inner primary wall of plant cells and the outer fibrous secondary wall with which he is chiefly preoccupied.

The beautiful work on the wall structure of green algae leading to the elucidation of cellulose chain direction is well described. The wall structure in the cells of flowering plants is discussed and a chapter is devoted to the primary wall of growing cells.

In a final chapter the mechanisms of orientation and growth are discussed. Particularly interesting are the speculations on the spatial relationships between cellulose and protein.

The book can be warmly recommended to all physicists interested in the impingement of their own discipline on botanical science. J. T. R.

The Polarizing Microscope, by A. F. Hallimond. 2nd Edn. Pp. 204. (York: Cooke, Troughton and Simms, 1953.) 15*s*.

We welcome this new edition of a monograph that already established itself in its first edition of 1948, and commend it without reserve to the serious user of the microscope. It is a valuable guide both to student and research worker and is beautifully illustrated. It is published by a maker of polarizing microscopes and, to some extent, partakes of the character both of a catalogue and a textbook. Indeed, the title of the book might more aptly have been 'The Cooke Polarizing Microscope', since the text is largely a guide to the use of the Cooke instrument. Nevertheless the possessor of any other polarizing microscope will also find this monograph a mine of valuable information, and an eminently practical guide, clearly written, by a very expert microscopist.

The field covered is wide and the theory included adequate. Valuable chapters are those on the determination of path difference, determination of extinction, the examination of minerals and the measurement of rotation. Considering the fact that a dozen pages have been, curiously enough, devoted to the description of polishing methods and hardness testing (this latter an aspect rather remote from the title), the omission of any description of the use of the Vickers Projection Microscope for polarization is indeed surprising for this large comprehensive instrument, made also by the maker of the Cooke polarizing microscope, is equipped with polaroids, and even if it is not strictly a polarizing microscope in the formal sense, it is often used as such.

Although this book is rather too specialized for the physics undergraduate, it should be ready to hand for every research worker who uses the microscope for revealing structural character: the physicist, geologist, metallurgist and biologist.

The reviewer noted only one minor misprint. On p. 41 there is a reference to a non-existent figure, figure 93. S. TOLANSKY.

Introduction to Electric Theory, by R. G. FOWLER. Pp. xii + 390. (Cambridge, Mass.: Addison–Wesley, 1953.) $7.50.

This book is designed as a text appropriate to certain mathematics and physics curricula of American universities. Its standard is roughly that of the pass degree of British universities. It is generally recognized that there is a great need for electricity textbooks of this kind. Simultaneously there is a great need for expositions which place theoretical electricity on a rigorous, logical footing and are free of the fallacies and pitfalls for which the subject is notorious. The ' tidying-up ' of classical electricity, however, still lies largely on the research level, and to attempt to combine this with a didactic account at the undergraduate level at the present time is a difficult task indeed. This is what the author has attempted. It is therefore sufficient commendation that he should have achieved a fair measure of success, and although much remains that one might criticize, no critic would wish to be unduly harsh—especially one who has himself had the experience of teaching such a course.

Regarding the treatment on its didactic merits, the outstanding characteristic is the unparalleled beauty, clarity and simplicity of exposition, making the subject not only understandable but alive and fascinating to the student grappling with its difficulties for the first time. Points which, as shown by experience, the student finds particularly difficult (e.g. the potential) are discussed at great length, leaving no room whatsoever for uncertainty or misunderstanding. Where possible, topics are illustrated by worked examples, and each chapter ends with a large number of examples for the student to do, with answers at the end of the book. This is in accordance with a plan of the author's to exploit the application to physics of the American student's parallel mathematical training, but no one will deny that its desirability is general. The text is well illustrated, a striking feature being the use of beautiful shaded diagrams for depicting complex, especially three-dimensional, situations. Vectors are introduced as a basic concept, but, appropriately so at this level, vector analysis is not used as a method of calculation; only vector summaries of results are given. A general system of units is used throughout, and is specialized where necessary, as in problems.

As already indicated, the scope of the book is roughly that of a pass degree course, so that a detailed description is unnecessary. The order differs from the conventional wherever this is necessary to meet the demands of an improved logical continuity. Features of special interest are the thorough use of Kelvin cavities for overcoming the difficulties of static field theory in polarized media, the introduction of convection and displacement currents at an early stage in order to provide a complete, although elementary, idea of the electromagnetic field, and the excellent and extensive treatments of special but useful topics such as networks and alternating current circuits.

Unfortunately the book could not serve as a complete text for the type of course usually given in British universities since, as indicated by the title, only the theory of electricity is treated; the experimental aspects are scarcely touched

upon. More unfortunate, perhaps, is the fact that the average student may not feel able to pay such a high price for a text dealing with a limited part of his course. However, in all other respects the book can be highly recommended to both teacher and student. A reviewer cannot allow for the individual tastes of teachers, but it is a book that the teacher must see and judge for himself.

R. CADE.

The Printing of Mathematics, by T. W. CHAUNDY, P. R. BARRETT and CHARLES BATEY. Pp. ix+105. (Oxford: University Press, 1954.) 15*s*.

Most physicists know that it is easier, and therefore cheaper, to print $(a+x)/(b+y)$ than $\frac{a+x}{b+y}$, and $\sqrt{(x+y)}$ than $\sqrt{x+y}$. They might be in doubt whether e^{ax_1} is any more difficult than e^{ax}, and most would be surprised to learn that x_1^2 is cheaper than x_1^2.

Here is a book which makes the reasons clear, by actually describing the processes of composing on a monotype machine and by hand. The authors are men of quite exceptional experience in the field of mathematical printing, and have written an account which cannot fail to interest anyone of enquiring mind; but they have gone further, and provide a reference work in which the author pressed for symbols may see what is available in different founts, whether a letter can easily be printed with a dot, bar, accent or tilde over it, and can even see a specimen legible handwriting for Latin and Greek letters.

It would have been great fun to mention a misprint, but this reviewer, at least, has failed to find one.

J. H. AWBERY.

Tables of 10^x (*Antilogarithms to the base* 10). NATIONAL BUREAU OF STANDARDS APPLIED MATHEMATICS SERIES, No. 27. Pp. viii+543. (Washington: U.S. Government Printing Office, 1953.) $3.50.

The main table gives values of 10^x to 10 decimal places, at intervals of 0·00001, for the complete range from 0 to 1.

There is a subsidiary table, by help of which it is possible to compute antilogarithms to 15 decimal places.

The table was compiled by correcting Dodson's table of 1742, and is printed from typescript. Differences are not tabulated.

J. H. AWBERY.

Physikalisches Wörterbuch, by W. H. WESTPHAL. Pp. v+795. (Berlin: Springer-Verlag, 1952.) D.M. 148.

When one reads in such an authoritative work as the *Oxford Dictionary* that an interference fringe is "one of a series of alternate light and dark bands produced by a diffraction grating", the need for specialist technical dictionaries is more than apparent. Moreover, a little thought is sufficient to show that a dictionary of physics must be radically different from a dictionary of words. To describe the adjectival use of 'Hamiltonian' by the phrase "pertaining to or invented by the Irish mathematician, Sir William Rowan Hamilton", is semantically impeccable so far as the conventional lexicographer is concerned, but the meanings that the word carries for the physicist are left unmentioned. More often than not the understanding of the meaning of a word or phrase in physics entails learning new facts and merely accepting the word or phrase as the accepted label for the given phenomenon.

In English there appears to be only the *Dictionary of Applied Physics*, and this is now very much out of date—but pure physics lacks even this. It is regrettable therefore that linguistic considerations may prevent many English-speaking physicists from possessing the volume under discussion. A person who knows the German word 'Ergodenhypothese', for example, is more than likely to know enough about statistical mechanics not to need to refer to a short dictionary article. However, a good English–German scientific dictionary used as an intermediary, not to mention the almost completely international word vocabulary of physics, should make the use of this *Physikalisches Wörterbuch* quite practicable.

The material of the book has been prepared by some eighty collaborators under the editorship of Professor Westphal. The list of contributors, all but three of whom are from Germany, contains a reassuring number of workers with international reputations. There are more than 1600 pages, of large format, and well printed in double columns on excellent paper. The number of entries exceeds 10 000, and the text contains 1595 figures. The material is very up to date, and an emphasis seems to have been given (unexpectedly) to theoretical and mathematical physics. Recent attempts to formulate a divergence-free quantum electrodynamics, for example, are described and references given.

The binding of the book is of a quality most of us long thought forgotten. A sample of entries revealed in most cases clarity and concise description. Every library, and many individuals, should possess a copy. H. H. HOPKINS.

Mechanics, by K. R. SYMON. Pp. xiii + 358. (Cambridge, Mass.: Addison–Wesley (Principles of Physics Series), 1953.) $7.50.

This book would provide a good basis in the mechanics of particles, rigid bodies and continuous media for the first two years of an honours physics course in this country. Judged from this point of view, it is admirably written and has few faults.

On most topics it is very thorough. Thus the theory and physical applications of the harmonic oscillator and of two coupled oscillators are given in great detail. Vector methods are used widely and this enables results to be obtained elegantly and in a general form, particularly the laws of fluid motion. The difficulties of interpretation of Newton's laws of motion and of the conservation laws are not hidden from the student and are well discussed.

This high standard fails to be maintained only when the temptation to add some parting information results in lack of thoroughness. Thus the sections on the Mach and Reynolds numbers, which end the chapter on continuous media, are not particularly clear. These and, perhaps, the section on electromagnetic forces might well have been omitted and room found instead to conclude the chapter on Lagrange's equations with the theory of small vibrations and normal modes. N. C. BARFORD.

CONTENTS OF SECTION A

PAGE

Mr. H. E. Hall. The Inertia of Heat Flow in Liquid Helium II 485

Mr. H. N. V. Temperley. A Possible Model of Liquid He_3 495

Dr. G. C. Fletcher. Calculations of the First Ferromagnetic Anisotropy Coefficient, Gyromagnetic Ratio and Spectroscopic Splitting Factor for Nickel . 505

Dr. G. L. Squires. A Nuclear Method for the Estimation of Traces of Light Water in Heavy Water 520

Dr. A. H. Cooke and Dr. H. J. Duffus. The Magnetic Susceptibility of Nitric Oxide in a Clathrate Compound 525

Dr. R. F. Barrow, Mr. J. A. T. Jacquest and Mr. E. W. Thompson. The Ultra-Violet Emission Spectra of the Gaseous Monofluorides of Gallium and Indium 528

Prof. D. R. Bates, Mr. R. T. S. Darling, Mr. S. C. Hawe and Dr. A. L. Stewart. Properties of the Hydrogen Molecular Ion: IV—Oscillator Strengths of the Transitions Connecting the Lowest Even and Lowest Odd σ-States with Higher σ-States 533

Dr. J. R. Prescott. The Decay of ^{207}Bi and the Energy Levels of ^{207}Pb . . 540

Research Notes:

Dr. M. Blackman and Mr. I. A. Khan. The Intensity of High Angle Kikuchi Bands 553

Prof. W. E. Burcham. The α-Activity Induced in Gold by Bombardment with Nitrogen Ions 555

Mr. D. A. Tidman. The Anomalous Scattering of μ Mesons 559

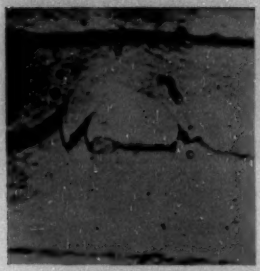
Letters to the Editor:

Miss D. G. Bell, Miss R. Hensman, Dr. D. P. Jenkins and Dr. L. Pincherle. A Note on the Band Structure of Silicon 562

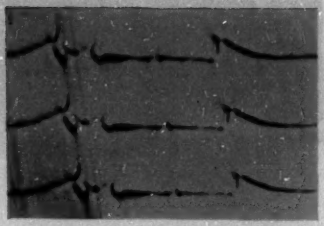
Dr. W. M. Fairbairn. The Deuteron Bombardment of Oxygen . . . 564

Reviews of Books 566

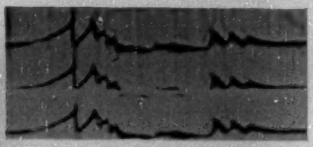
Contents of Section B 568



A
×160

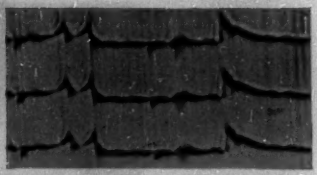


B
×185



C
×185

D
×185

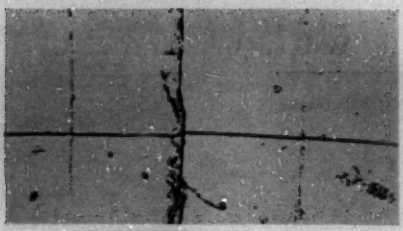


E
×185

F
×185

G
×32·5

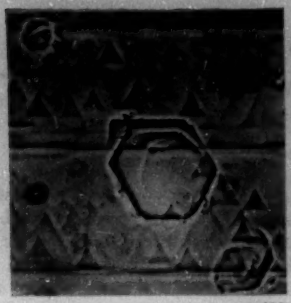
PLATE I.



A
×850

B
×215

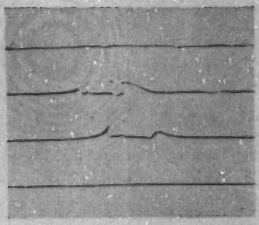
PLATE II.



(*a*) ×600

(*b*) ×800

(*c*) ×600



(*d*) ×225

(*e*) ×287

(*f*) ×400

(*g*) ×400

PLATE I.

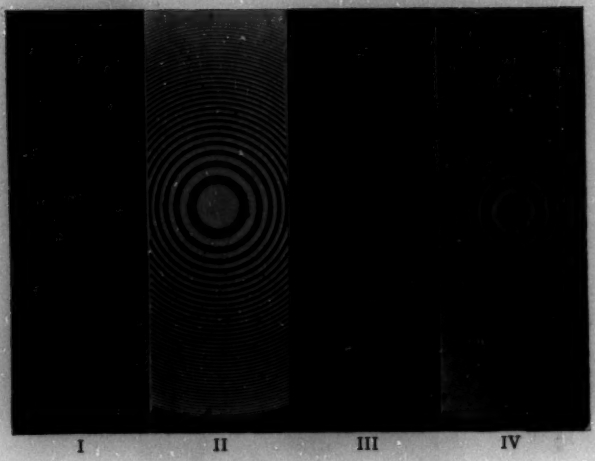


Figure 2.

Additivity of Colour Equations : II

By P. W. TREZONA*
Technical Optics Section, Imperial College, London

MS. received 4th December 1953, and in amended form 8th March 1954

Abstract. If c_1 units of one colour are matched by r_1, g_1 and b_1 units of red, green and blue respectively and c_2 units of another colour by r_2, g_2, b_2 units, then according to the law of additivity of colour equations the combined colour may be matched by r_1+r_2, g_1+g_2 and b_1+b_2 units respectively. Following an investigation into the validity of the law described in Part I of this paper, further experiments were designed, some to extend the problem and others to attempt to elucidate the cause of the small deviations which occur. Amongst the former was an experiment involving spectrophotometry. All experimental work was performed on the Wright colorimeter.

Additivity of colour equations is implicit in all colorimetric calculations. It is also of interest in relation to the visual processes.

Deviations were small enough to be caused by matches differing from their 'correct' value by less than a just-discriminable colour step. Some deviations were explained by a displacement within a colour step caused by a breakdown in Weber's law when applied to matching stimuli. This also explained the large deviations associated with poor discrimination.

§ 1. Introduction

The law of additivity of colour equations can be stated as follows: if c_1 units of a colour C_1 are matched by r_1, g_1 and b_1 units of the three matching stimuli R, G and B respectively, and if c_2 units of another colour C_2 are matched by r_2, g_2 and b_2 units of the same three matching stimuli, then an additive mixture of c_1 units of C_1 and c_2 units of C_2 is matched by r_1+r_2, g_1+g_2 and b_1+b_2 units of the matching stimuli. Algebraically

$$\text{if} \qquad c_1C_1 = r_1R + g_1G + b_1B$$
$$\text{and} \qquad c_2C_2 = r_2R + g_2G + b_2B$$
$$\text{then} \qquad c_1C_1 + c_2C_2 = (r_1+r_2)R + (g_1+g_2)G + (b_1+b_2)B.$$

The validity of this law was investigated (Trezona 1953, to be referred to as I) by choosing pairs of stimuli and matching them separately and in combination; this gave a means of comparing the sum with the mixture for each of R, G and B. Experimental work was performed on the Wright colorimeter (Wright 1946, p. 47), where R, G and B were 0·65 μ, 0·53 μ and 0·46 μ respectively. Results were expressed as percentage deviations from additivity, a positive deviation being defined as one where the sum was larger than the mixture. Results of a significance test at the 1% level were given. The passing of the test was denoted by $\surd$ and the failure by $\times$; $\times^+$ and $\times^-$ distinguished between 'positive' and 'negative' failures. The position of the sum relative to that of the mixture was shown on the chromaticity chart.

* Now at Newnham College, Cambridge.

The observations were of two types, A and B. In type A neither of the two colours combined was the same as that of a colorimeter matching stimulus. In type B a spectral colour was matched with four different amounts of a desaturating stimulus (D.S.) of the same wavelength as one of the colorimeter matching stimuli. Whereas type A gave three results, one for each of R, G and B, type B gave three results all for a matching stimulus, say R. These were R_1, R_2 and R_3, the result for the greatest amount of D.S. being R_3.

Although on the whole deviations are small, they appear to exist in a definite form. The present work extends the ground covered by I and also describes experiments designed to elucidate the nature of the deviations.

§ 2. Extension of the Range of The Problem

(i) *Extended Colour Combinations*

In I all colours chosen for testing additivity were spectral colours and each pair matched separately and in combination was complementary. Two variations were now made: (*a*) a pair of desaturated complementary colours was tested, (*b*) two spectral colours of fairly similar wavelength were used.

In (*a*) a yellow and a blue-green, both slightly desaturated, were employed and no D.S. was necessary. The blue-green colour was a mixture of $0{\cdot}55\,\mu$ and $0{\cdot}46\,\mu$, both from the test colour spectrum. The yellow colour was produced by a yellow filter in front of a lamp, the light being incident on a magnesium oxide sectored disc which presented the yellow and the blue-green to the eye alternately (Wright 1946, p. 58). The sector was rotated above the critical flicker frequency and, as before, the colours could be seen separately and in combination. The observer was P.W.T., who is protanomalous and usually gives large deviations in the red matching stimulus. The results (table 1) show that with desaturated colours deviations are still present but are rather smaller than with spectral colours.

Table 1. Comparison of Spectral and Desaturated Colours

	Matching stimuli					
	$0{\cdot}49\,\mu+0{\cdot}59\,\mu$			Desaturated yellow and blue-green		
	R	G	B	R	G	B
Percentage deviations	+ 8	0	0	+13	+4	0
	+11	−1	− 9	+16	0	−6
	+26	−1	− 8	− 4	0	+1
	+21	−2	−11	+14	+3	+1
Significance at 1% level	$\times^+$	√	$\times^-$	$\times^+$	√	√

Tests with a green desaturating stimulus

	$0{\cdot}42\,\mu+0{\cdot}53\,\mu$		
	G_1	G_2	G_3
Percentage deviations	+ 1	+13	+ 6
	+12	+15	+12
Significance at 1% level	$\times^+$	$\times^+$	$\times^+$

The two colours of fairly similar chromaticities used for (*b*) were $0{\cdot}51\,\mu$ and $0{\cdot}55\,\mu$ and the observer was P.W.T. For each matching stimulus there was no significant difference even at the 10% level.

In type B of the original investigation a test of additivity was made in turn on the red and blue matching stimuli. No test had been made using G because the spectral colours with which it combines are the violet and the extreme red, and in both of these regions discrimination is poor. For the sake of completeness, however, a few readings were taken for $0{\cdot}42\mu + 0{\cdot}53\mu$ by I.G.H.I. (table 1). A comparison with tables 2 and 3 of I will show that the same type of deviations occurs for G as for R and B.

(ii) *Change of Experimental Conditions*

(*a*) *Change of luminance level.*

A test was made on the maintenance of a colour match when the luminance level is changed. As was pointed out by König (1903), this is, in effect, testing the special case of additivity when the two colours added are the same.

König found that decreasing the luminance level over a considerable range had no effect on a colour match, but at very low intensities the match no longer held. The colours in which he found the variation were:

1. $0{\cdot}67\mu + 0{\cdot}52\mu$ against $0{\cdot}58\mu + 0{\cdot}475\mu$,
2. $0{\cdot}67\mu + 0{\cdot}59\mu$ against $0{\cdot}63\mu$.

Hering, however, showed that this change of colour match at low intensities does not occur if a small field is used or if the eye is kept light-adapted with a large field. Both these effects are consistent with the view that it is the inclusion of rod vision which causes changes in the colour match.

The field of the Wright colorimeter is 1° 20′ and covers only the rod-free area of the retina. Tests were made by two observers on König's colour combinations, the intensity range varying from the threshold to 100 times threshold. No variation of colour match could be detected. This was consistent with Hering's view. Other colour matches viewed in a similar manner were those made on each of the three colours used in testing the additivity of $0{\cdot}49\mu + 0{\cdot}59\mu$ in I. The intensity was changed from its level during the original experiment to one ten times lower, and again no change could be observed.

This result that a colour combines additively with itself is to be expected when two similar but not identical colours combine additively (§ 2(i)).

(*b*) *Change of adaptation.*

The colour matches made in testing additivity were done with the dark-adapted eye. Experiments to determine the effect of adaptation, both white and coloured, on a colour match are important in order to see if a test of additivity can be applied to an adaptation level other than that at which it was measured. If colour matches do not vary with the state of adaptation, then neither does additivity. If they do change, additivity may or may not hold for a given state of adaptation.

A colour match was made with the dark-adapted eye and then viewed when the eye was adapted to red, green, blue, yellow, blue-green and white stimuli (Wright 1946, p. 53). The adapting light was looked at for several minutes; then the test field was viewed for about 1 second and the adaptation field for 9 seconds alternately until it could be judged whether or not the match still held. The matches tested were the three used for the investigation on

$0{\cdot}49\mu + 0{\cdot}59\mu$. In every case the match still held even though the actual colour of the bipartite field altered considerably.

Wright (1936) found that although a colour match persists over a considerable range of adaptation intensities, both white and coloured, it breaks down at very high levels, above 15 000 photons. On a 2° field he found that at these high levels the R : G ratio required to match a yellow colour was very much increased. Moreover, the spectral coefficient curves were changed, indicating that the effect was due to a change in shape of the spectral response curves and not to changes of the macular pigmentation or any other pigment in the ocular media.

(iii) *An Experiment involving Spectrophotometry*

A common colorimetric procedure is to find the position of a colour on the C.I.E. chromaticity chart from a knowledge of its spectral transmission or reflection curve. For different wavelengths the transmission curve is multiplied by the relative energy curve of the source and by each of the C.I.E. distribution coefficients; then the relative areas under the X, Y and Z curves give the chromaticity coordinates. The multiplying process assumes additivity of similar colour equations and the integrating process assumes additivity of a large number of different colour equations.

The validity of this procedure can be tested by using the individual R, G and B spectral mixture curves instead of the X, Y and Z curves of the C.I.E. system. Then the calculated trichromatic coefficients can be compared with those obtained by a direct match. This was done by P.W.T., for the source S_A alone, for a yellow filter in front of the source and then for a blue-green filter. For the direct matches light was allowed to fall on a magnesium oxide surface. P.W.T.'s spectral mixture curve and the spectral transmission curves of the filters were measured on the Wright colorimeter, the latter by using it as a spectrophotometer (Wright 1946, p. 56). The positions of the directly matched and the calculated values on the Wright chromaticity chart (Wright 1946, p. 121) are shown in figure 1.

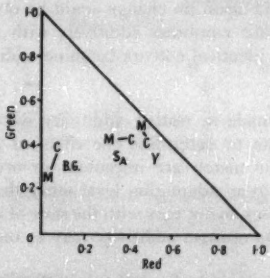


Figure 1. Positions on the Wright chromaticity chart: S_A, S_A alone; Y, S_A with yellow filter; BG, S_A with blue-green filter. M matched, C calculated. Observer: P.W.T.

Ishak (1951) made a comparison between the matched and calculated values of the white points S_A, S_B and S_C for three observers and of S_B alone for a

fourth observer (figure 2). In many cases the direct match is rather more blue than the calculated value.

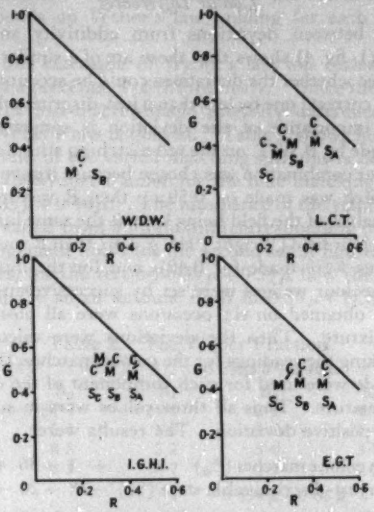


Figure 2. Positions of the white points on the Wright chromaticity chart for four observers (Ishak). M is the position of the match and C of the calculated colour.

(iv) *Additivity of Luminances*

Although the additivity of luminances in heterochromatic photometry is largely outside the scope of this work, the two are very closely linked. Abney (1913) used red, green and violet stimuli on which he made heterochromatic luminance matches separately and combined in various ways. He concluded that the luminances of these stimuli combined additively. More recently H. König (1948) showed that additivity held to within 2% for most colours and to within 5% for the remainder. Unfortunately in this type of observation the different methods of flicker and direct matching do not always give the same results.

Macadam's observations (1950) suggest that although deviations may be small they can be significant. A white mixture was kept unchanged in the comparison half of a field and the red component of an equally bright white mixture was reduced in intensity; the resulting colour was a bluish-green, very obviously *brighter* than the white. Thus by removing red light a brighter colour was produced. Two observers tried and failed to see this effect on the Wright colorimeter. In each case reduction of the red component made that field darker.

A special case of additivity of luminances arises when both halves of the field are of the same colour. Under these circumstances, if additivity of colour equations holds then so does additivity of luminances, although the reverse is not necessarily true.

§ 3. Experiments Designed for the Elucidation of the Problem

(i) *Individual Colour Matches compared with Just-Discriminable Colour Differences*

A comparison between deviations from additivity and just-discriminable colour differences (I, fig. 4) shows that these are of a similar order of magnitude. It is of interest to see whether the deviations could be accounted for by each match differing from the 'correct' one by less than a just-discriminable colour difference; if this is so, the importance of the deviation is somewhat diminished. An experiment was made by P.W.T. on the red matching stimulus for $0{\cdot}49\mu + 0{\cdot}59\mu$; this particular colour combination was chosen because it gave very large deviations for R. First a match was made on $0{\cdot}59\mu$; then R was made just perceptibly too large, the two halves of the field being kept at the same luminosity by a neutral wedge in the test-colour field (Wright 1946, p. 59); then R was made just too small. Similar observations were made for $0{\cdot}49\mu$ and for the mixture. In each case the 'red' and test-colour wedges were set by successive approximation.

The deviations obtained on six occasions were all positive, the sum being larger than the mixture. Then the deviations were calculated in a different way. Instead of taking the readings for the colour matches, the just-discriminable steps on the low side were used for each component of the sum and that on the high side for the mixture. Thus all three values were in such a direction as to tend to nullify the positive deviation. The results were:

Deviations based on colour matches (%)	+ 1	+36	+33	+15	+18	+27
Deviations based on just-discriminable steps (%)	−28	−26	−37	−49	−24	−51

Thus it is apparent that each match could be displaced from the 'correct' match by much less than a just-discriminable colour step and still give the observed deviation from additivity. Similar results were obtained by P.W.T. for the blue matching stimulus for $0{\cdot}5825\mu + 0{\cdot}46\mu$.

Although deviations are consistent with matches being 'displaced' within a just-discriminable step, the deviations are themselves visible. P.W.T. was shown the mixture of $0{\cdot}49\mu + 0{\cdot}59\mu$ together with the R, G and B quantities corresponding to the sum for each of six cases. Every time the upper half of the field, corresponding to the sum, was judged to be redder. This was in spite of the fact that these settings were presented at random, interspersed with the actual mixture settings. Thus the sum must have differed from the mixture by more than a just-discriminable step.

(ii) *Weber's Fraction*

It follows from Weber's law (I, p. 551) that $\Delta(\log I)$ is constant; hence a just-discriminable difference in the logarithm of the intensity should be independent of the intensity. Weber's law cannot be applied to a colour match where there are three matching stimuli instead of one without further testing. In making a colour match the procedure is something approaching the bracketing technique where each matching stimulus in turn is made alternately too large and too small, finally being left in a position half way between. Now for each matching stimulus the wedge which controls the intensity is linear with respect to $\log I$. So if a just-discriminable difference $\Delta(\log I)$ is the same on each side of the 'correct' match, the final setting, which is half way between the two

extremes, will be at the 'correct' position. If, however, $\Delta(\log I)$ is greater below the match than above it, the 'correct' position will be higher than the experimental reading half way between. Thus the validity of the colour-matching technique of 'bracketing' depends on Weber's law holding for each matching stimulus separately.

The validity of this extension of Weber's law was tested by expressing the standard deviation as a percentage of the intensity. Since the standard deviation is proportional to the just-discriminable intensity difference, the fraction will remain constant if Weber's law is obeyed. Table 2 shows how Weber's fraction increases as the amount of the varied matching stimulus is decreased. The first set of figures shows the Weber fraction for the blue matching stimulus B ($0{\cdot}46\,\mu$) in the match of $0{\cdot}5825\,\mu + 0{\cdot}46\,\mu$. The luminance of the yellow was kept constant while that of $0{\cdot}46\,\mu$ varied from a minimum (corresponding to B_0) to a maximum B_3. A similar arrangement holds for the second set of figures where R is $0{\cdot}65\,\mu$. The table also shows, for the test on $0{\cdot}49\,\mu + 0{\cdot}59\,\mu$ how the small amount of R in $0{\cdot}49\,\mu$ + D.S. and the small amount of B in $0{\cdot}59\,\mu$ + D.S. both give a large Weber fraction.

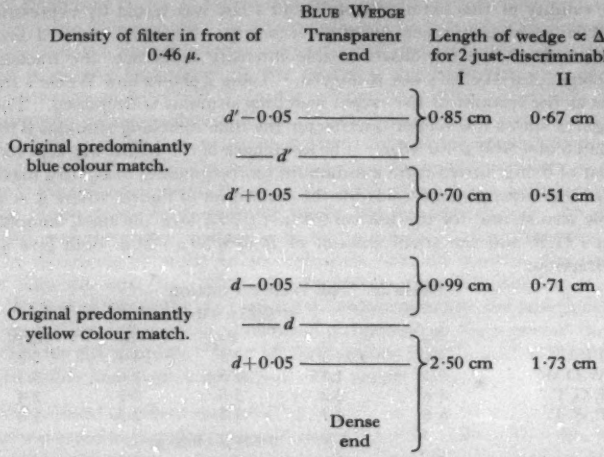
Table 2. The Weber Fraction

			$0{\cdot}5825\,\mu + 0{\cdot}46\,\mu$		
	B_0	B_1	B_2	B_3	B_0/B_3
J.E.G.	34·1	14·1	2·1	4·4	7·8
W.D.W.	5·1	1·8	1·8	1·6	3·3
E.G.T.	8·8	3·2	5·6	3·7	2·4
P.W.T.	8·5	4·7	3·1	2·2	3·8
			$0{\cdot}494\,\mu + 0{\cdot}65\,\mu$		
	R_0	R_1	R_2	R_3	R_0/R_3
W.D.W.	5·5	1·0	0·9	2·4	2·3
P.W.T.	13·6	3·2	4·0	4·3	3·2

W.D.W.	$0{\cdot}49\,\mu + 0{\cdot}59\,\mu$	$0{\cdot}49\,\mu$ + D.S.	$0{\cdot}59\,\mu$ + D.S.
R	0·7	*9·7*	1·5
G	1·0	0·5	1·6
B	1·7	2·4	*11·1*

These results show the general trend in the Weber fraction changes. An experiment was next performed to show how the just-perceptible step differed above and below a match. This is more directly related to the position of a match within a region of no discrimination. P.W.T. first took a yellow colour consisting of $0{\cdot}5825\,\mu +$ a little $0{\cdot}46\,\mu$ ($0{\cdot}46\,\mu$ had in front of it a fairly dense filter of density d) and matched it against R, G and B. Then the filter in front of $0{\cdot}46\,\mu$ was changed to $d - 0{\cdot}05$; R and G remained as they were, but B was altered, together with a neutral wedge over the whole T.C. field for equating luminosities. B was made just perceptibly too large in the absence of a luminosity difference and then just too small. The difference between the two scale readings gave, in units proportional to $\Delta(\log I)$, two just-discriminable steps when the amount of B used was greater than that in the original match. It was necessary to work with two just-discriminable steps, one on each side of $0{\cdot}5825\,\mu + 0{\cdot}46\,\mu$ (with $d - 0{\cdot}05$), in order to avoid basing a reading on the match of this colour. Next the filter in front of $0{\cdot}46\,\mu$ was changed to $d + 0{\cdot}05$ and two just-discriminable steps were found in the same way, this time the amount of B being less than that in the original match. All the above is an investigation on an original colour match of a predominantly

yellow colour. The procedure was repeated for a predominantly blue colour when the filter in front of 0·46 μ had a very much lower density d'. Thus we have quantities proportional to $\Delta(\log I)$ for values of B above and below a predominantly yellow colour and above and below a blue colour. The values obtained are represented in a way to show the whole procedure diagrammatically.



The differences between the figures in the upper part of the diagram are probably not significant. But for the predominantly yellow colour the just-discriminable step is very much larger on the lower side of the match than on the upper.

Thus, when there is a relatively large amount of blue in the mixture, discrimination is roughly equal on each side of the match and the 'correct' match for B will be half-way between the extremes of bracketing. But when the blue is present in much smaller quantities than the other matching stimuli, the discrimination step on the lower side of the match is much greater than that on the upper. Hence the 'correct' match is much further from the lower extreme of bracketing than from the upper, and the experimental reading lying half-way between will be too low.

§ 4. Discussion

Probably the most noteworthy results found in the whole observation are:

(i) Significant results appear to be associated in part with poor discrimination. One example of this is P.W.T.'s test for R on 0·49 μ + 0·59 μ, which was significant even at the 0·01% level (I, p. 553).

(ii) There is a general negative tendency for type A, the sum being smaller than the mixture. This occurs for the blue matching stimulus, to a lesser extent for the red and not at all for the green (I, tables 1 and 3).

(iii) In type B there is a general positive tendency which is more pronounced for B than for R (I, tables 2 and 3). This is not so great for small amounts of D.S. as for larger. In fact J.E.G. shows a consistently negative deviation for B_1 while B_2 and B_3 are positive (I, table 2).

In type A it is the red and blue matching stimuli that show negative deviations and not the green. In one of the matches the amount of R might be expected to be inaccurate because the predominant colour is blue-green. In another, a predominantly yellow colour, the amount of blue will be expected to be inaccurate, but G will be expected never to have this inaccuracy. The large values of Weber's fraction corresponding to the inaccurate R and B are shown in italics in table 2. In § 3 (ii) it was shown that for a yellow colour, when the amount of B is small, the 'correct' match will be higher than the experimental one which is half-way between the extremes of bracketing. Consequently the amount of B for one component of the sum will be too small, whereas it will be normal for the mixture and the other component of the sum as the colour is not predominantly yellow and Weber's law will apply. A negative deviation will result. This explains the negative deviations of type A and also those of type B when the amount of D.S. is small. Furthermore, the poorer the discrimination the larger the just-discriminable step and the greater the difference between the 'correct' and experimental amounts of B in the yellow match. Hence the bias and the resulting deviation which occur for normal people will be exaggerated for colour defectives.

A much more difficult problem is why positive deviations occur for type B. The effect is not associated with alterations in Weber's fraction; it must be large enough to outweigh the Weber negative effect. It is difficult to see how it could have an operational cause, and seems more likely that it is physiological. It must be considered in conjunction with the fact that changes of luminance level and adaptation do not affect a colour match. No comparatively simple explanation such as a non-linear relation between light input and photo-chemical decomposition, or even a change of shape of a photo-chemical absorption curve with change of concentration, will explain the two facts together. Some electrophysiology by Granit (1949) which might have some bearing on the problem showed that when two stimuli acted on the same retinal element under certain conditions they combined, while under others one stimulus tended to "behave as though non-existent".

The Significance of Deviations from Additivity

A consideration of deviations from additivity led to a discussion of Weber's fraction and hence to the effect of this on the colour match itself. It will be seen that, in making a colour match on a spectral colour, if a small amount of D.S. is used the 'correct' match will not be obtained. If, on the other hand, a very large amount of D.S. is used, inaccuracies will arise due to the subtraction of two large quantities to give a small one. The optimum amount is probably a little larger than that used in this work.

Among the colorimetric data affected by additivity are the spectral coefficient curves which give the proportions of R, G and B required to match wavelengths throughout the spectrum. Each of the three curves has a negative portion which is obtained by the use of a D.S.; hence they are dependent on the above effect. The spectral mixture curves involve multiplying the spectral coefficient curves throughout the spectrum by the appropriate height of the V_λ curve. Thus the spectral mixture curves depend on additivity in so far as the spectral coefficient curves depend on it. Also, the former assume that a colour match does not change with the luminance level; this has been shown to be true. Transformation

of the spectral coefficient data to another set of coordinates involves additivity since it includes the addition of three colour equations.

On the C.I.E. chromaticity chart (I, fig. 4) the effects of non-additivity are less important than on any individual chart, since a certain tolerance has to be allowed, in any case, because a given observer does not make colour matches in exactly the same way as the standard observer.

§ 5. Conclusion

Conclusions (i)–(vi) are given in I.

(vii) A desaturated blue–green combined with a desaturated yellow show slightly less extreme deviations than $0{\cdot}49\mu + 0{\cdot}59\mu$.

(viii) A change in luminance gives no change in a colour match. This can be considered as the addition of identical colours; addition of fairly similar colours also gives no deviation. A change in a colour match can occur at very low luminance level, but only when rod vision is included.

(ix) Change in adaptation, white and coloured, has no effect on a colour match and hence no effect on additivity. Wright found that the colour match breaks down under glare conditions.

(x) When a colour is matched directly and obtained from integrating the spectral mixture curves there is a slight difference. The match often appears bluer.

(xi) Deviations can be accounted for by matches being displaced from their 'correct' value within a just-discriminable step. However, in certain circumstances the sum and mixture can be seen to be different.

(xii) A consideration of Weber's fraction and the just-discriminable step gave an explanation of the negative deviations and the increased deviations associated with poor discrimination. There is no obvious explanation for the positive effect.

(xiii) Deviations from additivity have an effect on all colorimetric measurements, including the spectral coefficient curves, the spectral mixture curves and colour transformations. The effect on the C.I.E. chromaticity chart is probably not serious.

Acknowledgments

The author is very grateful to Professor W. D. Wright for suggesting the problem, for many helpful suggestions and criticisms and for acting as an observer. Thanks are also due to Dr. L. C. Thomson for his helpful suggestions and for acting as an observer, and to Mrs. E. G. Wassef, Dr. J. E. Gibbs, Dr. I. G. H. Ishak and Mr. W. S. Yang for observing and recording readings. The author gratefully acknowledges the financial support of the Medical Research Council.

References

Abney, W. de W., 1913, *Researches in Colour Vision and the Trichromatic Theory* (London: Longmans Green), p. 105.

Granit, R., 1949, *Acta physiol. scand.*, **18**, 281.

Ishak, I. G. H., 1951, *J. Physiol.*, **115**, 25.

König, A., 1903, *Gesammelte Abhandlugen zur Physiologischen Optik* (Leipzig: Barth), p. 110.

König, H., 1948, *Rev. Opt.*, **27**, 551.

Macadam, D. L., 1950, *J. Opt. Soc. Amer.*, **40**, 591.

Trezona, P. W., 1953, *Proc. Phys. Soc.* B, **66**, 548.

Wright, W. D., 1936, *J. Physiol.*, **87**, 23; 1946, *Researches on Normal and Defective Colour Vision* (London: Kimpton).

Third-Order Aberrations of 'Doubly Symmetric' Systems

By J. C. BURFOOT
Department of Natural Philosophy, University of Aberdeen

MS. received 18*th January* 1954

Abstract. The third-order geometric aberrations of non-rotationally symmetric imaging systems are classified, and their characteristics given diagrammatically for the case of a circular exit aperture. The systems are 'doubly symmetric', that is, they retain symmetry about each of two mutually perpendicular planes. Eleven new aberrations occur, three of which are third-order distortions, three give curved image surfaces, three are analogous to coma, and two to spherical aberration.

§ 1. Introduction

The third-order geometric aberrations of optical or electrostatic imaging systems characterized by rotational symmetry are usually classified as distortion, curvature of field, astigmatism, coma and spherical aberration. The classification is convenient analytically, and although practical systems do not usually exhibit one only of these aberrations, one of them is frequently predominant, and it is sometimes possible to determine this by inspection of the image. In electron optics it may become necessary to abandon rotational symmetry in order to correct certain of these aberrations (Scherzer 1947, Burfoot 1953), so that it is of interest to classify the new aberrations which arise. The traditional classification cannot be extended if no symmetry restrictions are imposed, and in fact the retention of symmetry about each of two mutually perpendicular planes* ensures that no second-order aberrations arise. (Second-order aberrations are absent in rotational symmetry.) This paper classifies the third-order aberrations of such 'doubly symmetric' systems, and presents their characteristics diagrammatically for the case of a circular exit aperture.

§ 2. The Third-Order Aberrations

Imaging properties cannot depend on the coordinate system chosen to represent them, so that in rotationally symmetric systems the aberrations must be expressible in terms of the rotational invariants. Individual rays are usually represented by the coordinates of their intersections with the object plane O and with the plane A of the exit aperture, beyond which no further deflections of rays occur. The rotational invariants in terms of these coordinates are $A=r_0^2$, $B=r_A^2$, $C=r_0 r_A \cos(\theta_A-\theta_0)$, $D=r_0 r_A \sin(\theta_A-\theta_0)$ in polar coordinates. The classification of the third-order aberrations is such (see, for example, Zworykin *et al.* 1945) that one aberration is associated with each second-degree combination of these four invariants, but the combination A^2 is independent of the aperture coordinates and produces zero aberration, while the number of aberrations is further reduced by one, because C^2+D^2 is identical with AB. Of the remaining eight combinations, three (AD, BD, CD)

* Their intersection forms the longitudinal optical axis of the system.

are dependent on the sign of $(\theta_A - \theta_O)$; these are the 'anisotropic' aberrations which occur only in the presence of magnetic fields, and which will not be further considered. AC, AB, C^2-D^2, BC, B^2 are associated respectively with third-order distortion, curvature of field, third-order astigmatism, coma, and spherical aberration.

In double symmetry the invariants (using cartesian coordinates $x=r\cos\theta$, $y=r\sin\theta$) are $a=x_O^2$, $b=y_O^2$, $c=x_A^2$, $d=y_A^2$, $e=x_Ox_A$, $f=y_Oy_A$, so twenty-one second-degree combinations occur, of which three are independent of x_A, y_A, and among which e^2 and f^2 are identical with ac and bd respectively. The remaining sixteen terms, in this form, do not include representations of the traditional aberrations. However, since rotational symmetry is a special case of double symmetry, it must be possible to reclassify them into five traditional and eleven new aberrations, and this may be done by taking the combinations of invariants as*

$(a+b)(e+f)$	$(a+b)(c+d)$	$(c+d)(e+f)$	$(c+d)^2$
$(a-b)(e-f)$	$(a+b)(c-d)$	$(c-d)(e-f)$	$(c-d)^2$
$(a-b)(e+f)$	$(a-b)(c+d)$	$(c-d)(e+f)$	c^2-d^2
$(a+b)(e-f)$	$(a-b)(c-d)+4ef$	$(c+d)(e-f)$	
	$(a-b)(c-d)-4ef$		

(In the second group $ef\equiv(abcd)^{1/2}$, so that the penultimate term is identical with C^2-D^2.) By the classical method (e.g. Zworykin *et al.* 1945) the coordinates δx, δy of an image point relative to the paraxial image point G† are proportional to the partial derivatives of these terms with respect to x_A and y_A respectively. The first group consists of a term representing third-order distortion, and three new terms which also image each point faithfully; the second group contains curvature and third-order astigmatism and three new aberrations producing curved image surfaces, two of which are astigmatisms; the third group is coma and three analogous new aberrations; and the last group is spherical aberration, and two analogous new aberrations which affect all object points equally.

§3. The Aberration Figures

Figure 1 gives the forms of the distortion terms. One of the new distortions would be difficult to distinguish by inspection from the classical 'pin-cushion' or 'barrel' distortions. Data for the remaining terms are given in table 1, together with suggested nomenclature.

The three astigmatisms form perpendicular line images in planes lkr_O^2 in front of, and lkr_O^2 behind, the paraxial image plane (henceforth referred to as P.I.P.), where l is the distance of that plane from the aperture plane. In traditional astigmatism, one of these lines (which one, depends on the sign of the appropriate k) is oriented parallel to the radius vector of the object point θ_O; in conjugate astigmatism, one is oriented at $-\theta_O$; and in right astigmatism

* The optical path length from (x_O, y_O) to (x_A, y_A), expressed as a polynomial, contains *inter alia* the terms $H_1ae+H_2bf+H_3af+H_4be$, for example. These may be written as

$$h_1(a+b)(e+f)+h_2(a-b)(e-f)+h_3(a-b)(e+f)+h_4(a+b)(e-f)$$

where $4h_1 \equiv H_1+H_2+H_3+H_4$ $\quad$ $4h_2 \equiv H_1-H_2+H_3-H_4$

$4h_3 \equiv H_1+H_2-H_3-H_4$ $\quad$ $4h_4 \equiv H_1-H_2-H_3+H_4$.

† It is assumed that first-order aberrations have been removed (Burfoot 1953).

the orientations are always in the symmetry planes of the system (hence the name suggested). The line centres lie on curved surfaces tangential to P.I.P. at its centre, and with curvature radius $M^2/2lk$ where M is the magnification of the system.

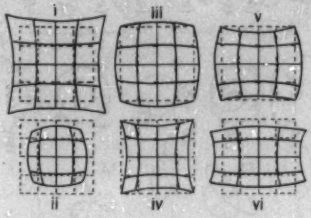


Figure 1. Third-order distortions.

The undistorted form is dotted.

(i) k_1 positive. Pincushion distortion.

(ii) k_1 negative. Barrel distortion.

$\delta x = k_1(x_0^3 + x_0y_0^2) \qquad \delta y = k_1(y_0^3 + x_0^2y_0)$

(iii) k_2 positive (iv) k_2 negative.

$\delta x = k_2(x_0^3 - x_0y_0^2) \qquad \delta y = k_2(y_0^3 - x_0^2y_0)$

(v) k_3 positive. Neg. k_3 rotates figure through 90°

$\delta x = k_3(x_0^3 - x_0y_0^2) \qquad \delta y = k_3(-y_0^3 + x_0^2y_0)$ } 'Hammock distortions'

(vi) k_4 positive. Neg. k_4 rotates figure through 90°

$\delta x = k_4(x_0^3 + x_0y_0^2) \qquad \delta y = k_4(-y_0^3 - x_0^2y_0)$ }

Curvature forms point images lkr_0^2 beyond P.I.P., i.e. on a similar curved surface with curvature radius $M^2/2lk$ on the axis. A negative k makes this surface lie in front of P.I.P. Cosine curvature forms point images $lk(x_0^2 - y_0^2)$ beyond P.I.P., i.e. on the surface sketched in figure 2.

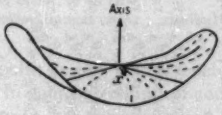


Figure 2. The image surface for cosine curvature.

Figure 3 illustrates for positive k in an exaggerated form the aberration figures in P.I.P. for the four comas, and also the figures to be observed in nearby planes. The full lines give the figures corresponding to fixed r_A (annular aperture), while the dotted lines indicate the envelope of all such figures up to the maximum value of r_A, set by the physical aperture, or otherwise (circular aperture). The heavy isolated point in each figure represents the principal ray, which in P.I.P. gives the paraxial image point G. The familiar coma 60° tail occurs of course because the 'annular' aberration figure does not enclose this point. It will be seen that the tail takes a larger angle than 60° in nearby planes, and disappears at more remote ones; similarly for conjugate coma. Table 2 gives some values for the tail angle. The elliptical comas do not exhibit tails,

Table 1

As a point in aperture plane describes a circle r_A about the optical axis, the corresponding point in paraxial image plane describes these figures relative to the paraxial image point G.

	$\delta x/k$ and $\delta y/k$	Type	Centre relative to G	Semi-axes or radius	*	Invariant
Curvature	$x_A r_O^2$, $y_A r_O^2$	Circle	0	$kr_A r_O^2$	θ_A	$(a+b)(c+d)$
Cosine curvature	$x_A r_O^2 \cos 2\theta_O$, $y_A r_O^2 \cos 2\theta_O$	Circle	0	$kr_A(x_O^2-y_O^2)$ or $kr_A r_O^2 \cos 2\theta_O$	θ_A	$(a-b)(c+d)$
Astigmatism	$r_A r_O^2 \cos(-\theta_A+2\theta_O)$, $r_A r_O^2 \sin(-\theta_A+2\theta_O)$	Circle	0	$kr_A r_O^2$	$-\theta_A+2\theta_O$	$(a-b)(c-d)+4ef$
Conjugate astigmatism	$r_A r_O^2 \cos(\theta_A+2\theta_O)$, $-r_A r_O^2 \sin(\theta_A+2\theta_O)$	Circle	0	$kr_A r_O^2$	$-\theta_A-2\theta_O$	$(a-b)(c-d)-4ef$
Right astigmatism	$x_A r_O^2$, $-y_A r_O^2$	Circle	0	$kr_A r_O^2$	$-\theta_A$	$(a+b)(c-d)$
Coma	$r_A^2(2x_O+x_O\cos 2\theta_A+y_O\sin 2\theta_A)$, $r_A^2(2y_O-y_O\cos 2\theta_A+x_O\sin 2\theta_A)$	Circle	$2kr_A^2x_O$, $2kr_A^2y_O$	$kr_A^2 r_O$	$2\theta_A-\theta_O$	$(c+d)(e+f)$
Conjugate coma	$r_A^2(2x_O+x_O\cos 2\theta_A-y_O\sin 2\theta_A)$, $r_A^2(-2y_O+y_O\cos 2\theta_A+x_O\sin 2\theta_A)$	Circle	$2kr_A^2x_O$, $-2kr_A^2y_O$	$kr_A^2 r_O$	$2\theta_A+\theta_O$	$(c+d)(e-f)$
Elliptical coma	$r_A^2(x_O+2x_O\cos 2\theta_A-y_O\sin 2\theta_A)$, $r_A^2(y_O-2y_O\cos 2\theta_A-x_O\sin 2\theta_A)$	Ellipse	$kr_A^2x_O$, $kr_A^2y_O$	$kr_A^2 r_O$, $2kr_A^2 r_O$	† (i)	$(c-d)(e-f)$
Conjugate elliptical coma	$r_A^2(x_O+2x_O\cos 2\theta_A+y_O\sin 2\theta_A)$, $r_A^2(-y_O+2y_O\cos 2\theta_A-x_O\sin 2\theta_A)$	Ellipse	$kr_A^2x_O$, $-kr_A^2y_O$	$kr_A^2 r_O$, $2kr_A^2 r_O$	† (ii)	$(c-d)(e+f)$
Spherical aberration	$x_A r_A^2$, $y_A r_A^2$	Circle	0	kr_A^2	θ_A	$(c+d)^2$
Rosette aberration	$x_A r_A^2 \cos 2\theta_A$, $-y_A r_A^2 \cos 2\theta_A$	Rosette	0	kr_A^2	$-\theta_A$	$(c-d)^2$
Star aberration	x_A^3, $-y_A^3$	Star	0	kr_A^3	$-\tan^{-1}\tan^3\theta_A$	c^2-d^2

For simplicity k has been taken as $2hl/\phi_1^{1/2}$ in nine lines of the table, but as $4hl/\phi_1^{1/2}$ in the last three lines. ϕ_1 is the potential in image space (e.g. Zworykin *et al.* 1945).

* Angular polar coordinate of any image point w.r.t. *figure centre*.

† Tan $\alpha = -\frac{1}{2}\tan 2\theta_A$ where α is the angular coordinate relative to the long semi-axes :
(i) Long axis oriented at $-\theta_O$. (ii) Long axis oriented at θ_O.

Table 2. Coma and Conjugate Coma

Angle of tail of image figure in various image planes (see Appendix).

$\lvert m/n\rvert$	0·5	0·75	1	1·5	2·5	∞ (P.I.P.)
Angle (deg.)	134	106	94	84	74	60

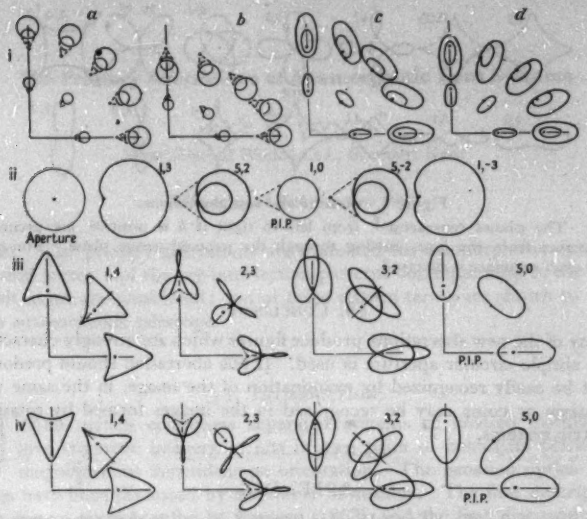


Figure 3. Third-order comas for various object points.
For details see text and Appendix. Arbitrary scales.

Paraxial image plane: (i) *a*, coma, *b*, conjugate coma, *c*, elliptical coma, *d*, conjugate elliptical coma.

Other planes:
(ii) coma and conjugate coma. These figures replace those in (i), in the same orientations. The planes represented, from left to right, are progressively further from the lens, passing through the paraxial image plane.
(iii) elliptical coma ; typical figures. The planes represented repeat in inverse order, as in (ii).
(iv) conjugate elliptical coma ; similarly.

even though the principal point is not always enclosed (e.g. figure 3, iii, *b*). Notice the word 'conjugate' in conjugate elliptical coma refers to the orientation of the ellipse centre relative to G, not the orientation of the major axis, which is the more striking property on inspection of the image. This nomenclature is suggested as a parallel with conjugate astigmatism, and with conjugate coma, which must be thus distinguished from traditional coma. The circular comas exhibit figures which differ only in scale and orientation for different object points. The figures due to the elliptical comas, however, differ in shape also, so that they have been represented only by a few typical figures (see Appendix). If k is negative, each of the individual shapes in figure 3 is turned through π about the appropriate principal point.

Figure 4 gives similar figures for rosette aberration and star aberration, which, like spherical aberration, affect each object point identically. The names describe the figure in P.I.P., but it will be seen that rosette aberration produces 'star-like' figures in other planes. They could easily be distinguished, however, since star aberration does not retain its four-fold symmetry as the plane is changed.

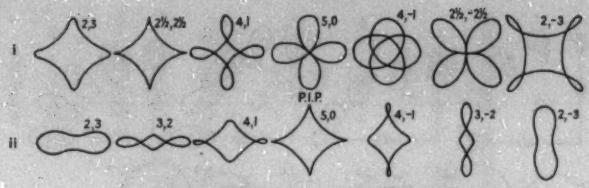


Figure 4. Rosette and star aberrations.

The planes represented, from left to right if k is positive, are progressively further from the lens, passing through the paraxial image plane. For negative k each sequence is reversed.

§4. Conclusion

Many of the new aberrations produce figures which are strongly characteristic when a simple circular aperture is used. If one aberration should predominate it could be easily recognized by examination of the image, in the same way as astigmatism or coma may be recognized in the images formed by rotationally symmetric systems.

APPENDIX

The figures for the elliptical comas (figure 3, iii and iv) have the general form

$$X = m(2\cos\theta_0\cos 2\theta_A \mp \sin\theta_0 \sin 2\theta_A) + n\cos\theta_A$$
$$Y = m(\mp 2\sin\theta_0\cos 2\theta_A - \cos\theta_0 \sin 2\theta_A) + n\sin\theta_A$$

where θ_A is the parameter, and the lower sign refers to conjugate elliptical coma. The m, n values are given on the figures. The principal ray passes at $(-m\cos\theta_0, \mp m\sin\theta_0)$ with respect to the X, Y origin (the figure 'centre').

The form of figures 3, ii is similarly

$$X = m\cos 2\theta_A + n\cos\theta_A$$
$$Y = m\sin 2\theta_A + n\sin\theta_A$$

with the principal ray at $(-2m, 0)$.

Figure 4 (i) is

$$X = (m\cos 2\theta_A + n)\cos\theta_A$$
$$Y = (-m\cos 2\theta_A + n)\sin\theta_A$$

and 4 (ii) is

$$X = (m\cos^2\theta_A + n)\cos\theta_A$$
$$Y = (-m\sin^2\theta_A + n)\sin\theta_A$$

with the principal ray at (0, 0).

In all cases the plane represented is distant nl/p from P.I.P. where $p \equiv n + m/kr_A r_0$ in figure 3 and $p \equiv n + m/kr_A^2$ in figure 4.

References

Burfoot, J. C., 1953, *Proc. Phys. Soc.* B, **66**, 775.

Scherzer, O., 1947, *Optik, Stuttgart*, **2**, 114.

Zworykin, V. K., Morton, G. A., Ramberg, E. G., Hillier, J., and Vance, A. W., 1945, *Electron Optics and the Electron Microscope* (New York: Wiley), Chap. 16.

The Primary Aberrations of Anamorphotic Lens Systems

By C. G. WYNNE
Wray (Optical Works) Ltd., Bromley, Kent

MS. received 29th January 1954

Abstract. The primary aberrations are evaluated for anamorphotic systems of cylindrical lenses, and the ray-intersection patterns corresponding to the sixteen different terms are considered; ten of these sixteen terms are shown to be zero for the anamorphotic telescope.

§ 1. Introduction

Under certain conditions separated systems of crossed cylinders can give stigmatic imagery at one or two pairs of conjugate planes, with magnifications dependent on orientation. The gaussian optics of such systems have been discussed by a number of authors. The first description of such a system appears to be by Farrenc (1862), and the first discussion of the theory by Abbe and Rudolph (1897); extensive references to the early literature are given by von Rohr (1899, 1920).

More recently Chrétien (1929, 1931) pointed out that stigmatic anamorphotic imagery could be achieved using a telescopic system of cylindrical lenses with their axes parallel. Such a system requires the addition of a normal lens system to yield a real image of a real object; it can be focused for different object distances by varying the cylinder separation to keep the image distance the same for the two principal sections, together with an appropriate focusing of the normal lens.

Chrétien, in one of his later publications (1931), discussed the design of such systems, stating, without proof, that they are free from primary aberrations if these are corrected for the principal section, using the normal Seidel formulae for a hypothetical system of spherical surfaces having the same curvatures, separations, etc. This is not the case, and it will be shown below that a number of additional primary aberrations arise for imagery out of the principal sections of any anamorphotic cylindrical system.

Chrétien does not discuss the primary chromatic aberrations of cylindrical systems, but it is shown below that these are corrected for a system for which they are corrected for spherical surfaced systems corresponding to the two principal sections.

Chrétien also gives—again without proof, but the proof is simple—an easy method of tracing rays through these systems, which may be summarized as follows. Taking rectangular coordinates with the z axis normal to all the surfaces and the x axis parallel to the cylinder axes, then if any ray make an angle $\mathcal{U}_x$ with the (y, z) plane in a space where the refractive index is n, then $n \sin \mathcal{U}_x$ is constant throughout the system; and the projection of the ray on the (y, z) plane

may be traced using the normal formulae for spherical surfaced systems but with a refractive index $n \cos \mathcal{U}_x$ in place of n. Chrétien gives this in a more complicated form which on a normal interpretation would become incorrect for cemented surfaces.

§ 2. Primary Aberrations of Cylindrical System

For a system of cylindrical refracting surfaces, whose axes are coplanar, take rectangular coordinates such that the cylinder axes lie in the (x, z) plane, normal to the z axis, which passes through the centre of the aperture stop. The aberrations along any ray from an oblique object point, relative to that along a central ray of the pencil, may be derived in terms of the paraxial data for rays in a principal plane section parallel to the (y, z) plane (for which the normal formulae for meridian section rays of spherical surfaces apply) and for rays in the (x, z) plane (for which the system is effectively a set of parallel plane surfaces). These will be called the first and second principal sections. Taking the origin on the individual surface to be considered, and suppose the object point for the first principal section lie at a distant η_y from the (x, z) plane, in a plane normal to the z axis distance l_y from the vertex of the surface, then for paraxial rays traced from this object point in a plane parallel to the (y, z) plane, at each surface

$$\Delta n \left(\frac{1}{r} - \frac{1}{l_y} \right) = 0$$

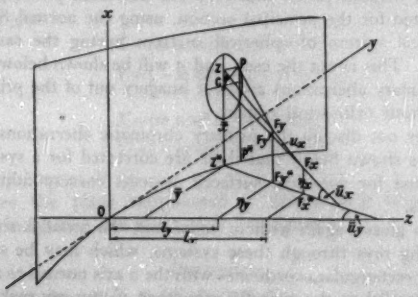
and throughout the system

$$n u_y \eta_y = H_y = \text{constant},$$

where Δ denotes the change on refraction, n is the refractive index of the space considered, r the radius of curvature of the principal section of the cylinder, and u_y the paraxial aperture angle of the pencil in the (y, z) plane.

Similarly, for a paraxial pencil in the (x, z) plane from an object point distant l_x from the pole of the surface, $\Delta n / l_x = 0$ and throughout the system $\eta_x = \text{constant}$.

Suppose now, at any surface of the system, C $(\bar{x}, \bar{y})$ be the point of incidence of the principal (central) ray of an oblique pencil (see figure); draw the line



CF_x through C, in such a direction that it cuts the plane $z = l_x$ at F_x distant η_x from the (y, z) plane, and also intersects the plane $z = l_y$ at a distance η_y from the

(x, z) plane, CF_x making an angle $\bar{u}_x$ with the (y, z) plane. Let P be the point of incidence on the surface of any ray of the pencil; and let Z be a point on the surface having the same x coordinate as P, lying on the generator through C. Join ZF_x, and let this intersect the plane $x=l_y$ at F_y : join PF_y. F_x, F_y are in the gaussian intermediate object planes for the surface in question, for imagery in the two principal sections of the cylindrical system; F_x and F_y will normally coincide in the initial and final spaces but not intermediately. There is, therefore, in general no intermediate focus with respect to which the primary aberrations at each surface of the whole wave front can be referred; but the primary aberration along a ray cutting the surface at P may be expressed relative to the ray through Z, with reference to a focus at F_y, and that of the ray through Z relative to the principal ray through C, with reference to a focus at F_x.

The derivations usually given of the primary (Seidel) aberrations for spherical surfaced systems break down either when the object lies at one of the refracting surfaces, or at one of their centres of curvature, a different expansion being required in the two cases. This is also true of the derivation which follows, which is not valid for the case of a field lens.

If C (and Z) lie at a distance $\bar{y}$ from the (x, z) plane and if Z^*, F_y^*, F_x^*, P^* be the projections of Z, F_y, F_x and P on the plane $x=0$, then to the order of approximation to be considered, if $\widehat{ZF_xC}=u_x$,

$$ZF_y = Z^*F_y^* \sec(u_x+\bar{u}_x) = Z^*F_y^*[1+\tfrac{1}{2}\sin^2(u_x+\bar{u}_x)+\ldots]$$

$$PF_y = P^*F_y^* \sec(u_x+\bar{u}_x) = P^*F_y^*[1+\tfrac{1}{2}\sin^2(u_x+\bar{u}_x)+\ldots],$$

and the aberration of the ray through P, relative to that at Z, referred to a focus at F_y is

$$\delta w_y = \Delta\{n(ZF_y - PF_y)\}$$

$$= \Delta\{n(Z^*F_y^* - P^*F_y^*)\} + \tfrac{1}{2}\Delta\{n[Z^*F_y^* - P^*F_y^*]\sin^2(u_x+\bar{u}_x)\}.$$

Now $\Delta\{n(Z^*F_y^* - P^*F_y^*)\}$ is the first order aberration for a pencil in the (y, z) plane of aperture Z^*P^* $(=ZP)$ and object size η_y, which is of precisely the same form as for a meridian pencil in a system of spherical surfaces of the same curvatures, etc. as the cylinders. For the second term, to the approximation required the angles u_x, $\bar{u}_x$ may be taken as equal to the corresponding angles for a paraxial ray traced in the (x, z) plane, and:

$$(Z^*F_y^*)^2 = \left(l_y - \frac{\bar{y}^2}{2r}\right)^2 + (\bar{y}-\eta_y)^2$$

$$= l_y^2\left[1 - \frac{\bar{y}^2}{l_y}\left(\frac{1}{r}-\frac{1}{l_y}\right) + \frac{\eta_y^2}{l_y^2} - 2\frac{\eta_y\bar{y}}{l_y^2}\right]$$

$$Z^*F_y^* = l_y - \tfrac{1}{2}\bar{y}^2\left(\frac{1}{r}-\frac{1}{l_y}\right) + \frac{1}{2}\frac{\eta_y^2}{l_y} - \frac{\eta_y\bar{y}}{l_y}$$

if $\bar{y}/l_y$ and η_y/l_y are small. Similarly

$$P^*F_y^* = l_y - \tfrac{1}{2}(\bar{y}+ZP)^2\left(\frac{1}{r}-\frac{1}{l_y}\right) + \frac{1}{2}\frac{\eta_y^2}{l_y} - \frac{\eta_y(\bar{y}+ZP)}{l_y}.$$

Now for a paraxial ray traced in the (y, z) plane with incident height y on the surface, $ny(r^{-1}-l_y^{-1}) = ni = A$, i being the angle of incidence of the ray; $y/l_y = u$, the angle the paraxial ray makes with the axis; and if the paraxial

principal ray in this plane, traced from an object of height η, makes an angle $\bar{\imath}$ with the surface considered, writing $n\bar{\imath} = \bar{A} = 1/y(H + A\bar{y})$

$$n(\mathrm{Z^*F_y^* - P^*F_y^*}) = \frac{1}{2}\left(\frac{\mathrm{ZP}}{y}\right)^2 Ay + \frac{\mathrm{ZP}}{y}\frac{\eta_y}{\eta}\bar{A}y$$

$$\tfrac{1}{2}\Delta\{n(\mathrm{Z^*F_y^* - P^*F_y^*})\sin^2(\bar{u}_z + u_z)\} = \tfrac{1}{2}n^2(u_z + \bar{u}_z)^2\left[\left(\frac{\mathrm{ZP}}{y}\right)^2 Ay + \frac{\mathrm{ZP}}{y}\frac{\eta_y}{\eta}\bar{A}y\right]\Delta\frac{1}{n^2}$$

to the approximation considered.

In general, the system will have a circular stop either in the object or the final image space, where the foci in the two principal sections coincide; if the paraxial ray in the (y, z) plane and the ray through P pass through the edge of this stop, then in its space $\mathrm{ZP}/y = \sin\phi_0$, ϕ_0 being the azimuth angle of P in this space; the azimuth ϕ of P will change through the system, but ZP/y will be invariant to the required approximation. In the same way, if the paraxial principal ray object height in the space containing the circular stop be $\eta = (\eta_y^2 + \eta_x^2)^{1/2}$, so that in this space $\eta_y/\eta = \sin\psi_0$, where ψ_0 is the azimuth angle of the object, η_y/η will remain constant through the system.

Similarly, if the paraxial ray in the (x, z) plane be traced so that it passes through the edge of the stop in the space where this is circular, making an angle U_x with the axis, then in this space $u_x/U_x = \cos\phi_0$, which is again constant through the system: and if in this plane the paraxial principal ray through an object of height η make an angle $\overline{U}_x$ with the axis, $\bar{u}_x/\overline{U}_x = \cos\psi_0$, which is again invariant.

The wave aberration required, for a ray through P relative to that through Z, referred to a focus at F_y is therefore

$$\begin{aligned}\delta w_y = {} & [\tfrac{1}{8}\sin^4\phi_0 S_{\mathrm{I}y} + \tfrac{1}{2}\sin^3\phi_0\sin\psi_0 S_{\mathrm{II}y} + \tfrac{1}{4}\sin^2\phi_0\sin^2\psi_0(3S_{\mathrm{III}y} + PH^2) \\ & + \tfrac{1}{2}\sin\phi_0\sin^3\psi_0 S_{\nabla y}] + n^2[U_x\cos\phi_0 + \overline{U}_x\cos\psi_0]^2[\tfrac{1}{4}Ay\sin^2\phi_0 \\ & + \tfrac{1}{2}\bar{A}y\sin\phi_0\sin\psi_0]\Delta\frac{1}{n^2} \qquad \ldots\ldots (1)\end{aligned}$$

where $S_{\mathrm{I}y}$, $S_{\mathrm{II}y}$, etc. are the normal Seidel coefficients for a spherical surfaced system corresponding to the first principal section of the cylinders.

The second component of the aberration of the oblique pencil to be considered is that of the ray at Z relative to that at C, referred to a focus at F_x.

To the order of magnitude to be considered

$$\mathrm{CF}_x = [(\mathrm{F}_x^*\mathrm{Z}^*)^2 + (\bar{x} - \eta_x)^2]^{1/2} = \mathrm{F}_x^*\mathrm{Z}^* + \frac{1}{2}\frac{(\bar{x} - \eta_x)^2}{\mathrm{F}_x^*\mathrm{Z}^*} - \frac{1}{8}\frac{(\bar{x} - \eta_x)^4}{(\mathrm{F}_x^*\mathrm{Z}^*)^3} + \ldots$$

$$\mathrm{ZF}_x = [(\mathrm{F}_x^*\mathrm{Z}_x)^2 + (\bar{x} - \eta_x + \mathrm{ZC})^2]^{1/2} = F_x^*\mathrm{Z}^* + \frac{1}{2}\frac{(\bar{x} - \eta_x + \mathrm{ZC})^2}{\mathrm{F}_x^*\mathrm{Z}^*} - \frac{1}{8}\frac{(\bar{x} - \eta_x + \mathrm{ZC})^4}{(\mathrm{F}_x^*\mathrm{Z}^*)^3},$$

where $\bar{x}$, η_x and ZC are assumed small compared with $\mathrm{F}_x^*\mathrm{Z}^*$.

$$\mathrm{CF}_x - \mathrm{ZF}_x = -\frac{1}{2}\frac{1}{\mathrm{F}_x^*\mathrm{Z}^*}[(\mathrm{ZC})^2 + 2(\mathrm{ZC})(\bar{x} - \eta_x)] + \tfrac{1}{8}\mathrm{F}_x^*\mathrm{Z}^*[(u_x + \bar{u}_x)^4 - u_x^4]$$

to the required approximation. Now

$$\mathrm{F}_x^*\mathrm{Z}^* = \left(l_x - \frac{\bar{y}^2}{2r}\right)\sec\bar{u}_y = l_x - \frac{\bar{y}^2}{2r} + \tfrac{1}{2}l_x\bar{u}_y^2$$

$$\frac{1}{\mathrm{F}_x^*\mathrm{Z}^*} = \frac{1}{l_x}\left[1 + \frac{1}{2}\frac{\bar{y}^2}{rl_x} - \tfrac{1}{2}\bar{u}_y^2\right]$$

where again $\bar{y}$ is assumed small compared with l_x and r. The required aberration of the wave front at Z, relative to that at C, is therefore

$$\delta w_x = \Delta n(\mathrm{CF}_x - \mathrm{ZF}_x)$$
$$= \frac{1}{4}\frac{\eta}{l_x}[(\mathrm{ZC})^2 + 2(\mathrm{ZC})(\bar{x} - \eta_x)]\Delta\left(\bar{u}_y{}^2 - \frac{\bar{y}^2}{rl_x}\right) + \tfrac{1}{8}n^4[(u_x + \bar{u}_x)^4 - u_x{}^4]\frac{l_x}{n}\Delta\frac{1}{n^2}.$$

The second term is the first order aberration for a pencil in the (x, z) plane, of the same form as the Seidel aberrations for a plane surface. Proceeding as before, this gives, after slight reduction of the first term,

$$\delta w_x = [\tfrac{1}{8}S_{\mathrm{I}x}\cos^4\phi_0 + \tfrac{1}{2}S_{\mathrm{II}x}\cos^3\phi_0\cos\psi_0 + \tfrac{3}{4}S_{\mathrm{III}x}\cos^2\phi_0\cos^2\psi_0$$
$$+ \tfrac{1}{2}\cos\phi_0\cos^3\psi_0 S_{\nabla x}] + \tfrac{1}{4}nl_xU_x\cos\phi_0[U_x\cos\phi_0$$
$$+ 2\bar{U}_x\cos\psi_0]\left[\bar{A}^2\Delta\frac{1}{n^2} - 2\bar{A}\bar{y}P + \frac{n\bar{y}^2}{l_x}P\right]\sin^2\psi_0 \qquad \ldots\ldots(2)$$

where $S_{\mathrm{I}x}$, $S_{\mathrm{II}x}$, etc. are the normal Seidel coefficients for plane surfaces, corresponding to the second principal section of the cylindrical system, and $\bar{y}$ now refers to the incident height on the surface considered of a paraxial principal ray in the (y, z) plane through an object of height η.

The two components of aberration in the two principal sections given in equations (1) and (2) are each additive surface by surface, and may be added together in a space where the foci F_x and F_y coincide. They are applicable directly to systems of parallel cylinders giving stigmatic imagery.

For a system of crossed cylinders, equations (1) and (2) give the aberrations of the first half, for which the x axis is taken parallel to the cylinder axes. The aberration at P relative to Z may now be referred to a focus at F_y', where CF_x cuts the plane $z = l_y$ instead of being referred to a focus at F_y, by the addition of a term $\frac{1}{2}n(l_x - l_y)u_x\bar{u}_x u_y{}^2$; the two reference points are then both on the principal ray through C. A similar treatment to that given above, but expressing the aberration at P relative to a point X on the generator through P and the principal section through C, and the aberration at X relative to C, gives two new equations which may be used for the crossed-cylinder system. This treatment will not be given here, since it is shown below that in general crossed cylinder systems appear unlikely to be of practical use.

The primary chromatic aberrations of a cylindrical surface follow directly from the formulae given above. Thus, for the portion of wave front at P relative to that at Z, the chromatic aberration arising at the surface is

$$(\delta w_y)_\lambda = \Delta\left[n(\mathrm{ZF}_y - \mathrm{PF}_y)\frac{\delta n}{n}\right]$$
$$= \frac{1}{2}\left(\frac{\mathrm{ZP}}{y}\right)^2 A y\Delta\left(\frac{\delta n}{n}\right) + \left(\frac{\mathrm{ZP}}{y}\right)\bar{A}y\Delta\left(\frac{\delta n}{n}\right),$$

by inserting the values given above, and neglecting powers of small quantities higher than the second, δn being the change of refractive index for the wavelength range considered. Similarly

$$(\delta w_x)_\lambda = \Delta\left[n(\mathrm{CF}_x - \mathrm{ZF}_x)\frac{\delta n}{n}\right]$$
$$= -\frac{1}{2}\left(\frac{\mathrm{ZC}}{y}\right)^2 ynu_x\Delta\left(\frac{\delta n}{n}\right) - \left(\frac{\mathrm{ZC}}{y}\right)yn\bar{u}_x\Delta\left(\frac{\delta n}{n}\right).$$

$(\delta w_y)_\lambda$ and $(\delta w_x)_\lambda$ correspond to the normal first order chromatic aberrations of the two principal sections (see for example Hopkins 1950), and there are no additional terms for rays out of the principal sections.

§ 3. Discussion of Results

The total wave aberration given by $\Sigma(\delta w_x + \delta w_y)$ comprises 16 terms; this is the total number of primary aberration terms determined by the symmetry of the system, and in the general case all these 16 terms appear in actual systems. The groups of these terms dependent on particular combinations of powers of the aperture and field will now be discussed separately, for the case discussed by Chrétien of systems of parallel cylinders; for crossed cylinder anamorphotic systems the results follow similarly, but are of greater complexity, each of the oblique aberration terms being doubled.

The ray intersection patterns on the focal plane may be derived from the wave aberration, for example using the equations given by Hopkins (1950) but for the general case of each aberration they give results of very considerable algebraic complexity; this may be the reason that the older methods of aberration theory, in terms of ray intersection distances, have not been applied to the problem of cylindrical lenses.

(*a*) *Spherical Aberration*

For a complete anamorphotic system, the aberration term depending on the fourth power of the aperture is, for a ray of aperture ρ at the circular pupil

$$\delta w = \left(\frac{\rho}{y}\right)^4 \left[\tfrac{1}{8}\sin^4\phi_0 \Sigma S_{Iy} + \tfrac{1}{8}\cos^4\phi_0 \Sigma S_{Ix} + \tfrac{1}{4}\sin^2\phi_0 \cos^2\phi_0 \Sigma\left(\pi^2 U_x^2 A y \Delta \frac{1}{n^2}\right)\right]. \quad \ldots\ldots(3)$$

The first two terms are respectively zero and a maximum at $\phi_0 = 0$ or π, and maximum and zero at $\phi_0 = \pi/2$ or $3\pi/2$; the third term is a maximum midway between these directions, at each of which it becomes zero. It is clearly always zero for an infinite object distance ($U_x = 0$), but not generally for a finite object distance. For $U_x = 0$, ΣS_{Ix} is also zero, so that for this special case correction of ΣS_{Iy} does give complete correction of axial aberration, in agreement with Chrétien's findings.

In this and the later cases, the ray intersection pattern with the focal plane will be referred to the object or image space in which the pupil is circular; in the other space the pattern will be the anamorphotic image of this.

When either ΣS_{Iy} or ΣS_{Ix} alone are finite, the other coefficients being zero, the ray intersection pattern is a straight line through the gaussian image, in the y or x direction respectively, the length of line, for an annulus of aperture of radius ρ, being proportional to ρ^3, and the distance along this line from the gaussian focus for the intersection point of a ray from this annulus of azimuth ϕ_0 varying as $\sin^3\phi_0$ or $\cos^3\phi_0$. When $\Sigma S_{Iy} = 0 = \Sigma S_{Ix}$, but the third term is finite, the ray intersection pattern in the focal plane of rays through a circular annulus of the pupil is given by the four-cusped hypotrochoid:

$$x = K\sin\phi\sin 2\phi, \qquad y = K\cos\phi\sin 2\phi$$

the scale factor K varying as the cube of the annulus diameter.

Equation (3) may clearly be divided in a number of different ways into terms dependent on different functions of ϕ_0. For example, writing equation (3) as:

$$\delta w = A\sin^4\phi_0 + B\cos^4\phi_0 + C\sin^2\phi_0\cos^2\phi_0,$$

this is equivalent to:

$$\delta w = A\sin^2\phi_0 + B\cos^2\phi_0 + D\sin^2\phi_0\cos^2\phi_0,$$

where $D = C - A - B$, or

$$\delta w = A + E\cos^4\phi_0 + F\sin^2\phi_0\cos^2\phi_0,$$

where $E = B - A$, $F = C - 2A$, and in each of these cases the ray intersection pattern for each term taken separately will be different; for the last equation, the first term A will correspond to the ordinary rotationally symmetrical spherical aberration of a spherical surfaced system corresponding to the first principal section of the cylinders. There is thus a certain arbitrariness in the discussion of the separate terms; but the form of the aberration coefficients given in equations (1) and (2), which is the one chosen here for analysis, seems the most significant since it segregates those terms depending wholly on the imagery characteristics of one of the two principal sections, and those dependent on both. This division of terms has therefore been chosen for each of the groups of aberration terms considered below.

(*b*) *Coma*

The terms depending on the cube of the aperture and the first power of the field are:

$$\delta w = \left(\frac{\rho}{y}\right)^3\left[\tfrac{1}{2}\sin^3\phi_0\sin\psi_0\Sigma S_{\mathrm{II}y} + \tfrac{1}{2}\cos^3\phi_0\cos\psi_0\Sigma S_{\mathrm{II}x} + \tfrac{1}{2}nU_x\sin\phi_0\cos\phi_0 \times \Sigma\{\sin\phi_0\cos\psi_0 n\bar{U}_x A + \cos\phi_0\sin\psi_0 nU_x\bar{A}\}\, y\Delta\frac{1}{n^2}\right].$$

For a given value of ψ_0, the first term is a maximum at $\phi_0 = \pi/2$ or $3\pi/2$ and zero at $\phi_0 = 0$ or π, the size of the maximum depending on ψ_0, the term being zero for all ϕ_0 at $\psi_0 = 0$ or π: for the second term similar considerations obviously apply. The third term is again zero for an infinite object distance ($U_x = 0$), as is the second, so that again correction of $\Sigma S_{\mathrm{II}y}$ gives complete coma correction for this special case.

When $\Sigma S_{\mathrm{II}y}$ is finite, the other aberrations being zero, the ray intersection pattern for a circular annulus in the pupil is a straight line in the y direction, with one end at the gaussian image, the length of line being proportioned to $\sin\psi_0$, the line being traversed to and fro twice as the annulus is circulated once. $\Sigma S_{\mathrm{II}x}$ alone gives a similar line in the x direction. For each of the two parts of the third term taken alone, the ray intersection pattern is an ellipse, whose minor axis is half its major axis in length, the gaussian focus lying at one end of the minor axis, which lies in the x direction for the first part of the third term, and the y direction for the second, the size of the ellipse for a given azimuth ψ_0 depending on ρ^3, and for a given radius of annulus depending on $\cos\psi_0$ and $\sin\psi_0$ respectively for the two parts of the term. For the two parts of the third term taken together, the ray intersection pattern is still an ellipse whose major axis is twice its minor axis, lying with its centre distant from the the gaussian focus by a distance equal to its semi-minor axis, but with an orientation depending in general both on ψ_0 and on the relative magnitudes of the two parts of the term. In each case the ellipse is traversed twice for a single rotation round the annulus.

(c) *Astigmatism and Field Curvature*

The terms depending on the square of the aperture and field are

$$\delta w = \frac{1}{4}\left(\frac{\rho}{y}\right)^2\Big[\sin^2\phi_0\sin^2\psi_0\Sigma(3S_{\mathrm{III}y}+PH^2)+\cos^2\phi_0\cos^2\psi_0\Sigma 3S_{\mathrm{III}x}$$
$$+\sin^2\phi_0\cos^2\psi_0 n^2\bar{U}_x{}^2\Sigma\left(Ay\Delta\frac{1}{n^2}\right)$$
$$+4\sin\phi_0\cos\phi_0\sin\psi_0\cos\psi_0 n^2U_x\bar{U}_x\Sigma\left(\bar{A}y\Delta\frac{1}{n^2}\right)$$
$$+\cos^2\phi_0\sin^2\psi_0 n^2U_x{}^2\Sigma\frac{l_x}{n}\left\{\bar{A}^2\Delta\frac{1}{n^2}-2\bar{A}\bar{y}P+\frac{n\bar{y}^2}{l_x}P\right\}\Big]. \qquad \ldots\ldots (4)$$

Here again the first two terms have maxima and minima with changing ϕ_0 in the same places as for the first two coma terms, the size of the maxima again varying with ψ_0. For the special case when $U_x=0$ the fourth term, and the last term become zero, as does $\Sigma S_{\mathrm{III}x}$, but the third term remains finite, giving in general a wave aberration:

$$\delta w_{(U_x=0)}=\frac{1}{4}\left(\frac{\rho}{y}\right)^2\Big[\sin^2\phi_0\sin^2\psi_0\Sigma(3S_{\mathrm{III}y}+PH^2)$$
$$+\sin^2\phi_0\cos^2\psi_0 n^2\bar{U}_x{}^2\Sigma\left(Ay\Delta\frac{1}{n^2}\right)\Big].$$

For this case there are therefore two focal lines for off-axis imagery, which lie in the x and y directions for all object positions, the line in the y direction, corresponding with fans of rays for which $\phi_0=0$ or π, lying always in the gaussian focal plane, and the line in the x direction, corresponding to fans for which $\phi_0=\pi/2$ or $3\pi/2$, lying on a surface of double curvature. If the primary aberration of the first principal section is zero $[\Sigma(3S_{\mathrm{III}y}+PH^2)=0]$ there remains in general an astigmatism giving a cylindrically curved field for fans of rays parallel to the (y, z) plane.

For the general case when all the terms in equation (4) are finite, the foci for focus of rays parallel to the two principal sections lie on two quadratic surfaces but due to the presence of the fourth term these do not correspond to the two principal sections of the wave front, and therefore do not give line foci parallel to the x and y directions, except when $\sin 4\psi_0=0$; for other object orientations there are two line foci at right angles to each other, lying on two quartic surfaces, at orientations depending on the five aberration coefficients as well as on ψ_0.

(d) *Distortion*

The terms depending on the first power of the aperture and the cube of the field are:

$$\delta w=\frac{1}{2}\left(\frac{\rho}{y}\right)\Big[\sin\phi_0\sin^3\psi_0\Sigma S_{\nabla y}+\cos\phi_0\cos^3\psi_0\Sigma S_{\nabla x}+\sin\phi_0\sin\psi_0\cos^2\psi_0 n^2\bar{U}_x{}^2$$
$$\times\Sigma\left(\bar{A}y\Delta\frac{1}{n^2}\right)+\cos\phi_0\sin^2\psi_0\cos\psi_0 n^2U_x\bar{U}_x\Sigma\frac{lx}{n}\left\{A\Delta\frac{1}{n^2}-2\bar{A}\bar{y}P+\frac{n\bar{y}^2}{l_x}P\right\}\Big].$$

Here again the second and fourth terms are zero for an infinite object, and the first if the primary aberrations of the first principal section be corrected, but in general the third term remains finite. Taken by itself, it represents a displacement of the actual image point from its gaussian image position in the y direction only; and is zero for object points on the x or y axes; for an object in the form of a

circle centred on the z axis, the image becomes an ellipse from the anamorphotic effect, the points of the ellipse on the major and minor axes being undistorted, and the maximum distortion occurring for the points on the circle of azimuth ψ_0 given by $\sin^2\psi_0=\frac{1}{3}$. For a number of applications of anamorphotic systems, including the one with which Chrétien was concerned, the distortion is of no importance, since the image is subsequently rectified by projection back through a similar system to the one that formed it.

In general, for an anamorphotic system of cylinders with a finite object or image distance ($U_z\neq 0$) sixteen conditions have to be satisfied to ensure correction of the primary monochromatic aberrations, in place of five for a spherical system. This will obviously require systems with a large number of surfaces in order to give enough variables, and where enough variables are available, the finding of solutions will be of great mathematical complexity. For an infinitely distant object and image, imaged anamorphotically by a system of parallel cylinders, ten of the sixteen conditions are automatically satisfied, leaving only six to be considered. Such a system may give a real object of a real image by the addition of normal spherical systems before and behind the anamorphotic ' telescope ', and clearly lends itself much more readily to aberration correction than do systems of crossed cylinders.

References

Abbe, E., and Rudolph, P., 1897, " Anamorphotisches Linsensystem " (Carl Zeiss, Jena), German Patent 99722.

Chrétien, H., 1929, British Patent Specification 282,078, 289,454; 1931, *Ibid.*, 356, 955; 351,927.

Farrenc, L., 1862, British Patent Specification, 1453/1862.

Hopkins, H. H., 1950, *The Wave Theory of Aberrations* (Oxford: University Press).

von Rohr, M., 1899, *Theorie und Geschichte des Photographischen Objectivs* (Berlin: Springer); 1920, *The Formation of Images in Optical Systems* (London: H.M. Stationery Office).

The Formation of Trigons on Diamonds

By A. HALPERIN*
Royal Holloway College, University of London, Englefield Green, Surrey

Communicated by S. Tolansky; MS. received 23rd February 1954

Abstract. The results of observations on trigons on octahedral faces of diamonds are given. Imperfect steps are believed to be present on the growing surfaces and to cause the trigons to be formed. These steps are assumed to be formed as a result of accumulation of interstitial faults in the lattice.

Further, it is shown that trigons are formed mainly below a certain temperature, characteristic to the growing face and to the imperfect steps on it. This temperature should be lower than the transition temperature for the same face on a perfect crystal, and seems to be thus reduced to temperatures at which the crystals might have been grown.

The theory accounts for the observed characteristics of the trigons.

§ 1. Introduction

Equilateral triangular pits, or trigons (Sutton 1928), are very common on the natural octahedral faces of diamonds. Recently, using newly developed techniques, it has been found that even the morphologically most perfect diamonds available have their octahedral faces covered by an enormous number of very shallow trigons (Omar 1953, Tolansky 1953; see also figure 1 in the present paper).

No satisfactory explanation of the formation of trigons has yet been given. For long they were in fact assumed to be etch figures (Miers 1902, Smith 1930). It was recognized, however, that they are very different from the etch figures on diamond, and it became evident that they are growth features (Fersmann and Goldsmidt 1911, Sutton 1928, Williams 1932, Tolansky and Wilcock 1946). As to the formation of trigons, no explanation which accounts for their characteristics was given. The aim of this paper is to give such a theory.

§ 2. Characteristics of the Trigons

The trigons appear as very regular equilateral triangles on the octahedral faces of the diamond. They are always oriented one way on untwinned faces, having their points directed to the edges of the octahedron. They are mostly shallow with their bottoms flat, especially on nearly perfect crystals. The trigons shown in figure 1 (Plate), for example, are mostly less than 50 Å deep. On imperfect crystal faces they are, however, much deeper, and occasionally their sides are extended down inward to meet in a point, like an inverted pyramid.

On the faces of many diamonds the trigons are distributed at random and are a variety of sizes (figure 2 (Plate)). On the other hand, on diamonds which show mainly clearly developed shallow growth pyramids the trigons tend to be nearly uniform in size, and are aranged in rows on the terraces along the edges of growth sheets (figure 1).

* On study leave; permanent address: Department of Physics, The Hebrew University, Jerusalem, Israel.

In striking distinction from the egdes of the growth sheets, which are usually kinked and curved, the edges of the trigons are perfectly straight. This basic fact was not accounted for in earlier attempts which have been made to describe the origin of the trigons.

Sutton (1928) considers that trigons are a result of growth at different rates, growth at accelerated rate forming hills and at retarded rate trigons. This statement does not explain any of the characteristics of the trigons, and no reasons are proposed for the existence of different rates of growth.

Van der Veen (1913) assumes that trigons have been formed by three growth sheets approaching from three sides. A sketch by Van der Veen illustrating

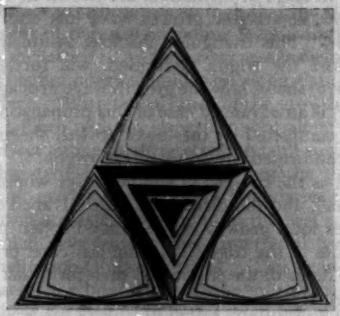


Figure 3. A sketch by Van der Veen (1913) showing his ideas on the formation of trigons.

this is reproduced in figure 3. In the figure we see the three growth hills built of thin sheets; the ' older ' ones, says Van der Veen, conform more nearly to the trigon and are straight, while the edges of the newer elementary plates are curved. No attempt was made to explain this descriptive observation.

More recently Tolansky and Wilcock (1946), while investigating interferometrically the octahedral faces of diamonds, propose similarly that trigons can be formed as a result of the arresting of three advancing plane growth-waves inclined at 60° to each other. While trigons *can* be formed in this way, this is not the case in general, as will be shown below. Besides, the characteristics of the trigons, and especially their perfect straight edges, remain unexplained.

§ 3. Some Observations on Trigons

In figure 1 part of an octahedral face of a highly optically perfect diamond crystal is shown. A shallow growth pyramid is seen emerging from a centre of growth. Edges of more growth pyramids are seen at the sides of the picture. The terraces left on the growth sheets are full of small trigons, directed opposite to the triangular growth sheets. (Experiments in this laboratory establish without any doubt that these shallow trigons are not due to etch.) The main direction in which growth advances is away from the centre and perpendicular to the edges of the growth sheets. Along each of these edges there is only *one main direction* of growth. When the growth wave approaches a trigon, the process is as follows: (i) in the beginning the growth wave approaches the edge of the

trigon which is parallel to the growth front, (ii) on reaching this edge (where the trigon should be formed) growth is arrested along it, (iii) growth proceeds on both sides of the obstacle, propagating round the corners and, finally, by inward growth, the sheets form the other sides of the trigon. When neighbouring trigons disturb the completion of a trigon from the side, growth propagates around all the obstacles and finally the trigons may be completed by growth backward.

These stages cannot of course be observed on completed trigons. Fortunately, many trigons on crystal faces are incomplete, being at different stages of formation when the growth stopped. Some of these incompleted trigons, making a series of trigons advancing in stage of formation, are shown in figure 4 (Plate).

Figure 4(*a*) shows the arrested growth wave just after reaching the edge of the *base of the trigon*. (This triangular obstacle, assumed to be formed on the growing surface before the trigon is formed, will further be called in short 'tribase'). Figure 4(*b*) shows how the growth wave propagates round the tribase corners. Figure 4(*c*) is an advanced case of this propagation, and in figure 4(*d*) the trigon is already completed by the lower growth sheets, and only the upper sheets have not yet conformed the edges of the trigon.

This is suggested as the process in simpler cases where the number of main growth centres is small, and the direction of the advance of the growth wave is clear. In other cases local growth centres may happen to start growing from two or even three sides of the trigon. In the last case, provided the directions of growth are clearly defined, the trigon can then form just in the way described by Van der Veen (figure 3).

Information concerning the formation of the tribases is obtained from figures 4(*a*), 4(*b*) and 4(*c*). In these figures we have the edge of the growing sheet on which the tribase should be formed quite near to the edge where growth was arrested by the obstacle. Geometrically there is no room for a full equilateral triangular obstacle on the surface. This shows that the formation of the obstacle itself takes place gradually with the advance of the growing surface on which this obstacle is formed. The first step is the formation of a very thin obstacle having a certain length. This advances with the growing sheet in the form of a truncated equilateral triangle which finally results in a full tribase.

§ 4. The Formation of Trigons

The obstacles which arrest the growth and cause the trigons to be formed are surely faults on the surface, faults in the arrangement of the atoms on the growing surface.

In recent years it has been shown that dislocations play an important part in the process of growth of crystals (Frank 1949, Burton *et al.* 1951). Perfect dislocations (producing identity translations, transferring atoms from one lattice point to another) are assumed to exist on a growing surface, thus providing the necessary perfect steps for continued growth at experimentally known rates. The predicted growth on the steps provided by dislocations was confirmed experimentally and growth spirals originating in screw dislocations were observed on many crystals (for references see Verma 1953).

Imperfect dislocations are defined as those which transfer atoms from lattice positions to twin-lattice positions (Frank 1949). Their strength is only a fraction

of the lattice unit, and the imperfect steps provided by these dislocations will not, usually, promote growth to such a degree as perfect ones.

It will have been assumed that under certain conditions of growth of diamond there is a high probability of imperfect steps being created on the octahedral growing surfaces, for example by creation of the sessile dislocation $(\frac{1}{3}a, \frac{1}{3}a, \frac{1}{3}a)$ suggested by Frank (1949) for face-centred cubic lattices.

The process of formation of these dislocations may be as follows: Interstitial faults are first formed by misfitting of individual atoms in the lattice points. This might be attributed to impurity atoms, but the possibility of the carbon atoms themselves forming such faults cannot be excluded (see discussion below). These faults accumulate from layer to layer in the growing crystal until their density on a certain surface area becomes so high that the surface energy may be lowered by stretching out of the new oncoming layer into the next mechanically stable configuration of the lattice. An imperfect sessile dislocation is thus formed on the boundary. This boundary should pass along the (111) twinning lattice planes cutting the surface. With the advance of the growth sheet this process is completed, leaving a tribase with an imperfect step along its boundary. It is separated from the perfect lattice by three stacking fault surfaces in the directions of the three octahedral twin-lattice planes cutting the growing face. These directions are known to predominate as twin boundaries in face-centred cubic lattices, for neighbouring atoms can match up in the best way across boundaries in these directions, which have a much lower interfacial free energy than boundaries of different orientation (Fullman 1951). This tendency should be more pronounced in the case of the boundaries of the imperfect steps on the growing diamond surface, as these are formed in better equilibrium conditions than those in which an annealing twin is formed. In this way the tribases are obtained, and in certain conditions of growth, the advancing layers will be arrested on the boundaries forming triangular growth pits or trigons.

§ 5. The Arresting of Growth by the Tribases

We shall follow Kossel (1927) by taking for demonstration a simplified model of the crystal, the molecules of which are cubes packed face to face (figure 5).

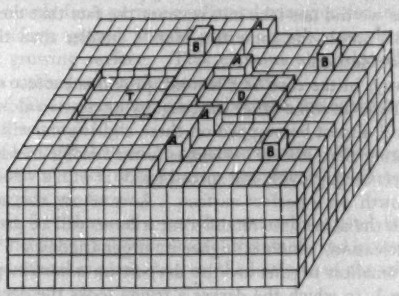


Figure 5. A Kossel (1927) crystal, modified by the addition of areas D, forming imperfect steps on the surface. The area T is analogous to the assumed tribases on the octahedral growing surfaces of diamond.

The figure shows an incompleted layer on a (001) surface of the crystal with its kinked step. The sites A are the kinks on the step and B are molecules which diffuse on the surface. Vacancy sites and pairs of adsorbed molecules etc. are omitted in the figure. The surface areas D form imperfect steps on the surface, their level being higher than the surrounding surface area by some fraction of the lattice unit. T was obtained from an area like D after the advancing growing layer, which was arrested on its edge, has surrounded it and conformed to its edges.

This arresting of growth means that the velocity of growth on this area is small compared with the velocity of growth on the surrounding surface. We shall see soon that this might be the case.

The velocity of growth from vapour (the case of growth from vapour is taken only for simplicity) on an advancing straight step (the radius of curvature of which is ∞) is given by the formula (Burton *et al.* 1951)

$$v_\infty = 2\sigma x_s \nu \exp(-W/\boldsymbol{k}T)\beta C_0$$

where σ is the supersaturation, x_s the diffusion distance on the surface, ν a factor of the order of the frequency of vibration of the molecules, W the evaporation energy, $\boldsymbol{k}$ Boltzmann's constant, T the absolute temperature, and β a constant usually smaller than unity. The factor which concerns us now is C_0, which is proportional to $1/x_0$, where x_0 is the mean distance between kinks along the step, given (Burton *et al.*) by $x_0 = \frac{1}{2}a\exp(w/\boldsymbol{k}T)$, where w is the energy of formation of a kink and a is the distance between two neighbouring molecules. For the perfect step sites w is small and for a close packed step on a (1, 1, 1) face of a face-centred cubic lattice it is estimated to be $w = \frac{1}{2}\phi$ where ϕ is the nearest neighbour interaction. On the other hand, for an imperfect step (on the edge of D, figure 5) formation of a kink means exchange between a perfect lattice site in the step and an imperfect site on the higher level. In this case w is much larger. Let us take it, for example, to be $w_d = 2\phi$, then we obtain for a typical value of $\phi/\boldsymbol{k}T \sim 4$

$$v_p/v_d \sim \exp(3\phi/2\boldsymbol{k}T) \sim 4 \times 10^2.$$

We see that the velocity of growth across the imperfect step is small compared with the velocity of growth in the surrounding perfect lattice. In fact v_d should even be smaller as we did not take into account the fact that the binding energy between the atoms across the imperfect step is smaller than that between the atoms in the perfect lattice.

In a similar way, in the case of diamond, provided perfect steps will not be formed on the tribases in some other way, e.g. by nucleation along an edge of the tribase, growth will almost be stopped on them and trigons will be formed.

It is worth noting that, once a nucleus is formed along an edge of the tribase across the imperfect step, growth will spread inside it with a velocity nearly equal to that of the growth on a perfect surface. As a result, the number of layers which will grow in the same time on different tribases will be proportional to the probability of nucleation, which is higher in larger trigons. This may explain the fact that the smallest trigons are the deepest on a certain part of a crystal surface (see figure 1, in which the darker a trigon looks the deeper it is). After a layer inside a tribase is completed, its level is higher than the surrounding lattice surface. The tasks are now exchanged, and growth may start from the tribase as a centre.

§ 6. Growth on Imperfect Steps

The probability of surface nucleation on the tribases as well as 'edge nucleation' (building up a nucleus along the boundary of a tribase) is surely a function of supersaturation and of temperature. The problem of surface nucleation in the case of a perfect lattice was undertaken theoretically by Burton *et al.* (1951 Parts III and IV). They have shown that at ordinary temperatures this probability is negligibly small unless the supersaturation is very high (of the order of 50%). For higher temperatures, above a certain 'transition temperature' T_c, the probability of surface nucleation increases and the surface becomes 'rough' even at low values of supersaturation. The exact calculation of T_c is complicated. By applying approximations it was shown, however, that T_c should be very high, and seems to be of the same order of magnitude or even higher than the melting point of the crystal.

These high transition temperatures are obtained when jumps of atoms are allowed only between levels of the perfect lattice. Things seem to be different for surfaces containing imperfect steps, where for the atoms on the boundary, intermediate levels, those of the imperfect steps, are allowed in addition to those of the perfect lattice. This should cause these edges to become 'rough' at a temperature much lower than the transition temperature for the perfect surface, and we may expect it to occur within the range of temperatures in which the crystals grow. This roughness of the boundaries provides the edge-nucleus for growth on the imperfect step.

We conclude, therefore, that at higher temperatures growth will take place on the whole surface, including the tribases, and almost no trigons will be formed. On the other hand, below a certain temperature, characteristic to the growing face and to the imperfect steps on it, growth will be stopped on the edges of the tribases and trigons will be formed.

§ 7. Discussion and Conclusions

The given theory accounts for the formation of the trigons and for their characteristics.

A short summary of conclusions is given now:

1. The arresting of growth on the boundary of the tribases can be explained.

2. The trigons mostly form perfect equilateral triangles, for in order to match best across the boundary, growth should pass along the twin-lattice octahedral planes cutting the growing surface. This explains also why all trigons point to the edges of the octahedron on the growing face.

3. The edges of the trigons are straight, because the energy of formation of kinks across the boundary is high.

4. Trigons can be very shallow on certain crystals (which appear therefore optically very perfcet) if these crystals grew mainly at higher temperatures, at which almost no trigons are formed.

5. The observed fact that, on a certain part of a crystal face, the deeper trigons are smaller than the shallower is also explained, and gives support to the assumption that growth takes place even on the bottoms of the trigons in certain conditions.

With crystals having rough faces things are more complicated. These faces contain very deep trigons, which consist usually of a large number of equilateral triangular boundaries, of which the outer ones are on higher levels on the surface.

When these combined trigons have a flat bottom, this bottom is often covered with an enormous number of tiny trigons (figure 6 (Plate)).

In many cases growth takes place inside such trigons, and it seems that such trigons were formed at comparatively high temperatures and supersaturation. Although the probability of edge nucleation is high in these conditions, a delay may be expected, and when growth in these conditions proceeds in thick layers it results in large terraced trigons with clustered growth sheets on the terraces of the largest of them.

Changes of pressure and supersaturation might also affect the formation of the trigons. It seems that reduction in pressure should enhance the faults in the lattice and hence also the formation of the tribases. Changes of supersaturation, besides affecting the rate of growth, should also affect the growth on the tribases, as higher supersaturation should enhance the edge nucleation.

The main assumption, on which this theory is based, is that of the accumulation of imperfections in the lattice to form the sessile dislocations. This assumption is justified by the results which fit the observed characteristics of the trigons. Further support has been obtained by an interesting case of slip on diamond observed at this laboratory (in preparation for publication).

In this, trigons are concentrated along the slip lines with their centres just on the slip lines. It is evident that the slip must have occurred *before* the growth of the diamond was stopped. The most probable explanation of the high concentration of trigons on the slip lines is that these trigons resulted from distortions in the lattice which took place while slipping. This fits the above assumption by which trigons should result from a distortion in the lattice. It seems that in this case the formation of the trigons cannot be attributed to foreign impurities; this gives some support to the notion that the tribases can be formed without 'activation' by impurities.

It is interesting to note that growth pits resembling diamond trigons, but hexagonal in shape, have been observed by A. R. Verma (1954) on octahedral faces of a single crystal of germanium.

It is well known that diamond behaves in a peculiar way with regard to many of its physical properties, and these peculiarities have been attributed to mosaicity in its structure. There are, surely, connections between the mosaicity of the surface given by the trigons and between the mosaic structure of the diamond crystal in bulk.

Acknowledgments

It is a pleasure for me to thank Professor S. Tolansky for his help and continuous interest and encouragement in the course of this work, and for the facilities of his laboratory given to me. Thanks are also due to Dr. F. C. Frank for helpful criticism.

I wish also to express my thanks to UNESCO for a scholarship awarded to me, and to the Hebrew University of Jerusalem for granting me study leave.

References

Burton, W. K., Cabrera, N., and Frank, F. C., 1951, *Phil. Trans. Roy. Soc.* A, **243**, 299.
Fersmann, A., and Goldsmidt, V., 1911, *Der Diamant* (Heidelberg: C. Winter).
Frank, F. C., 1949 a, *Discussions of the Faraday Society*, No. 5; 1949 b, *Proc. Phys. Soc.* A, **62**, 202.

FULLMAN, R. L., 1951, *J. Appl. Phys.*, **22**, 448.
KOSSEL, W., 1927, *Nachr. Ges. Wiss. Göttingen*, 135.
MIERS, H. E., 1902, *Mineralogy* (London : Macmillan).
OMAR, M., 1953, *Thesis*, University of London.
SMITH, G. F. H., 1930, *Gem Stones* (London : Methuen).
SUTTON, J. R., 1928, *Diamond* (London : Thomas Murby).
TOLANSKY, S., 1953, *Research, Lond.*, **6**, 8.
TOLANSKY, S., and WILCOCK, W. L., 1946, *Nature, Lond.*, **157**, 583.
VAN DER VEEN, A. L. W. E., 1913, *Z. Kristallogr.*, **51**, 545.
VERMA, A. R., 1953, *Crystal Growth and Dislocations*) (London : Butterworths Scientific Publications) ; 1954, *Proc. Phys. Soc.* B, **67**, 359.
WILLIAMS, A. F., 1953, *The Genesis of the Diamond* (London : Ernest Benn Ltd.).

Cathode Sputtering in Glow Discharges

By G. ECKER* and K. G. EMELÉUS
Physics Department, Queen's University, Belfast

MS. received 11th February 1954

Abstract. The analysis and interpretation of experiments on cathode sputtering in glow discharges requires definite knowledge of the behaviour of the metal atoms after they leave the cathode. Previous theoretical investigations of this neglect important factors, or are based on imperfectly realizable assumptions, and are not in good agreement with experiment. In particular, application of these theories leads to qualitative and order-of-magnitude discrepancies with low-pressure experiments. We have now attempted to make a more precise description of the behaviour of the metal atoms. It has been shown that hitherto largely neglected effects of initial scattering, lateral boundaries, distribution of current over the cathode, high emission velocities and the difference in mass between metal atoms and gas molecules, are of fundamental importance. The new formulation removes most of the old difficulties and discrepancies. It also shows definitely, however, that the phenomena are so complex that it will probably never be possible to develop a theory of sufficient accuracy to permit of exact conclusions being drawn from the experimental data.

§ 1. Introduction

The observational material about sputtering of cathodes of glow discharges is extensive although, unfortunately, the bulk of the older work is unsuitable for analysis because the conditions under which the experiments were performed were incompletely specified. More recently experiments have been done under somewhat better controlled conditions, in particular by Guntherschulze and Blechschmidt. These have been summarized by Glockler and Lind (1939) and by Massey and Burhop (1952).

Three distinct problems are involved: (i) the atomic mechanics of the collision processes which liberate metal atoms from the cathode surface; (ii) the movement of the sputtered atoms to the internal boundaries; (iii) the process of deposition of sputtered atoms.

The present paper deals primarily with the second of these, under conditions when the gas pressure is sufficiently high for the motion of the atoms to be governed, broadly speaking, by diffusion through the main gas in the discharge tube. Under these conditions knowledge of (ii) is necessary for solutions of (i) and (iii). Partial solutions of this problem (ii) have been given by, amongst others, von Hippel (1926), Seeliger (1942) and Townes (1944), but in all cases approximations have been made and important factors ignored. This restricts severely the range of application of the results. It is therefore not surprising that serious discrepancies have been found between the results of theory and experiment (Guntherschulze 1942, Seeliger 1942, Massey and Burhop 1952).

* On leave of absence from Institut für Theoretische Physik, University of Bonn.

The conclusions drawn from the present work are that:

(*a*) these discrepancies probably arise from neglect of one or more of the following factors: (1) the difference of mass between metal atoms and gas molecules, (2) high velocity of ejection of metal atoms from the cathode, and their scattering in the gas, (3) radial variation in the discharge parameters;

(*b*) the use of sputtering experiments of the type discussed to obtain information about the mechanism of disruption of the cathode surface is likely to be very limited even when allowance has been made for these factors.

§ 2. Theory of the Movement of the Sputtered Atoms

To obtain results which may be discussed and tested fairly readily we shall consider a common practical form of discharge tube, consisting of a cylindrical vessel of diameter $2R$ containing two metals discs as anode and cathode distance d apart, each of diameter $2R$, with its plane perpendicular to the axis of the tube. For the discharges we are considering, the mean free paths of the metal atoms are small compared with R and d, and the partial pressure of the metal vapour is small compared with the gas pressure. Let z denote distance from the cathode measured towards the anode, and r radial distance from the axis of the tube. It will be assumed that the discharge is symmetrical about the axis. If this assumption is not made, the subsequent analysis becomes more involved, without appreciable gain in understanding of the essential features of the problem.

A certain number of metal atoms j will be liberated per cm^2 per sec from the cathode surface. Since j depends on the ion current density, which varies over the cathode surface (Childs 1930, Chaudhri and Baqui 1952), it will depend on r and be given by

$$j = j_0 f(r/R) \qquad \ldots\ldots(1)$$

where the form of the function f can be left unspecified.

Baum (1927) has shown that the energy with which the metal atoms are ejected may be several times greater than in the gas at a greater distance from the cathode. The passage of metal atoms into the gas cannot be described exactly by any boundary condition of diffusion theory, but must be considered as a scattering process. The zone in which the majority of the metal atoms undergo their first collision may be called the initial zone, and will extend outward for a distance z_a. z_a must be chosen so that the number of metal atoms which pass through the initial zone without making a collision is small compared with the effective cathode loss, i.e. the difference between the number of atoms leaving the cathode and those diffusing back. It will appear later that the exact value of z_a is not of importance. After their first collisions, provided persistence is allowed for in the scattering process, the metal atoms diffuse through the gas. The concentration n of metal atoms which have made at least one collision will be governed by a conservation law. The formulation of this is simplified by the fact that we can restrict the thickness of the initial zone to a few free paths. Within this zone there will be a rapid variation of n with z, and hence we can treat the problem approximately as unidimensional. The conservation equation then becomes

$$-\frac{\partial}{\partial z} D \frac{\partial n}{\partial z} = \frac{j}{l} e^{-z/l} \qquad \ldots\ldots(2)$$

where l is the mean free path of the primary metal atoms and D the diffusion coefficient.

Two points must be noticed about l and D. The first is that l will be different from the free path λ of a metal atom in thermal equilibrium with the gas, given by

$$\lambda = \frac{1}{n_g Q}\left(\frac{m}{M+m}\right)^{1/2}\frac{1}{1-P} \qquad \ldots\ldots(3)$$

$$P = (M-\alpha m)/(M+m) \qquad \ldots\ldots(4)$$

where n_g is the concentration of gas molecules, Q a collision cross section between metal atoms and gas molecules, and α a function of m/M whose value is between 0 and 1/3; M and m are the masses of the metal atoms and gas molecules respectively.

Let us put

$$l = K\lambda \qquad \ldots\ldots(5)$$

and assume that the distribution of velocities amongst the atoms on emission can be approximated to by a Maxwell law with mean velocity v_0. v_0 will be in general different from v_M, the mean velocity corresponding to thermal equilibrium with the gas. K will then be given by

$$K = \frac{2}{3}\left\{\left(1+\frac{M}{m}\right)\Big/\left(1+\frac{v_M{}^2 M}{v_0{}^2 m}\right)\right\}^{1/2}. \qquad \ldots\ldots(6)$$

The factor 2/3 allows for the fact that the emission directions follow approximately a cosine law (Seeliger and Sommermeyer 1935). The remainder of equation (6) expresses the dependence of free paths on the mean relative velocity.

Secondly, D must be a function of z for the following reasons. The mean velocity of the diffusing metal atoms immediately in front of the cathode can be considerably greater than v_M, because a considerable proportion of the atoms have then undergone few collisions since they left the cathode, and consequently have velocities of up to the order of the emission velocities. The increased mean velocity increases the mean free path. Both effects give an increase of the diffusion coefficient immediately in front of the cathode. An exact description of the dependence of D on z is not possible. It is certain, however, that D falls in a few free paths to nearly the value D_0 which it possesses in equilibrium. We will therefore write

$$D = D_0(1+D'e^{-z/l}) \qquad \ldots\ldots(7)$$

where

$$D' = (D_c - D_0)/D_0. \qquad \ldots\ldots(8)$$

D_c is the diffusion constant immediately in front of the cathode, and

$$D_0 = \lambda v_M/3. \qquad \ldots\ldots(9)$$

Equation (2) can now be integrated, giving

$$D_0(1+D'e^{-z/l})\frac{\partial n}{\partial z} = je^{-z/l} + a_1 \qquad \ldots\ldots(10)$$

and

$$n = -\frac{j\lambda K}{D_0 D'}\ln(1+D'e^{-z/l}) + \frac{a_1}{D_0}\{z+\lambda K\ln(1+D'e^{-z/l})\} + a_2 \qquad \ldots\ldots(11)$$

where a_1 and a_2 are constants independent of z but dependent on r, to be found from boundary conditions.

Beyond the initial zone (z_a) the concentration of the metal atoms will be governed by the equation

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial n}{\partial r} + \frac{\partial^2 n}{\partial z^2} = 0. \qquad \ldots\ldots(12)$$

This region will be referred to as the diffusion zone.

Since n is finite at $r=0$, the general solution of (12) can be written as a linear combination of expressions

$$n_\nu = \{c_{1,\nu} \exp(s_\nu z) + c_{2,\nu} \exp(-s_\nu z)\} J_0(s_\nu r) \qquad \ldots\ldots(13)$$

where the separation parameters s_ν and the constants $c_{1,\nu}$ and $c_{2,\nu}$ are determined by boundary conditions.

At the walls and electrodes

$$nv/4 = -D(\mathbf{i} \cdot \operatorname{grad} n) \qquad \ldots\ldots(14)$$

where $\mathbf{i}$ is the outward directed normal unit vector of a surface element. This boundary condition involves the assumption that a metal atom sticks at first impact. Although this is not always true (Ditchburn 1933), neglect of partial reflection does not alter our general conclusions.

Applying (13) and (14) to the cylindrical walls determines s_ν from the relations

$$J_0(s_\nu R)/\{s_\nu R J_1(s_\nu R)\} = 4\lambda/3R. \qquad \ldots\ldots(15)$$

Since $\lambda \ll R$, to the first order in λ/R

$$s_\nu = A_\nu \left(1 - \frac{4\lambda}{3R}\right) \Big/ R \qquad \ldots\ldots(16)$$

where A_ν is the νth zero of J_0, and hence the approximation

$$s_\nu = A_\nu/R \qquad \ldots\ldots(17)$$

is adequate.

The constants $c_{1,\nu}$, $c_{2,\nu}$, a_1 and a_2 are found by applying (14) to cathode ($z=0$) and anode ($z=d$), and from the conditions that n and D grad n shall be continuous at the boundary of the initial and diffusion zones ($z=z_a$). For this purpose we develop $f(r/R)$ (equation 1) in a Bessel–Fourier series

$$f(r/R) = \sum_i f_i J_0(A_i r/R). \qquad \ldots\ldots(18)$$

The analysis is straightforward. With the previous restrictions z_a does not appear explicitly in the result, which is, for the diffusion zone,

$$n = \frac{4j_0\epsilon}{v_M} \sum_\nu f_\nu J_0(A_\nu r/R) \frac{\sinh\{A_\nu(d-z)/R\}}{\sinh\{A_\nu d/R\}} \qquad \ldots\ldots(19)$$

$$\epsilon = \frac{D_0}{D_c} + \frac{3}{4}\frac{K}{D'} \ln(1+D'). \qquad \ldots\ldots(20)$$

The disappearance of z_a is satisfactory, as it has no precise physical significance. The formula for the initial zone will not be given, because it does not lead to results which are readily visualizable. Graphs of (11) and (19) are given together in the figure. They show, amongst other features, that n has a maximum value at a certain distance from the cathode.

From the above analysis we can derive expressions for the amount of metal leaving any unit area of the cathode which does not diffuse back (j_e), and for the total amount of metal leaving the cathode which does not diffuse back (F).

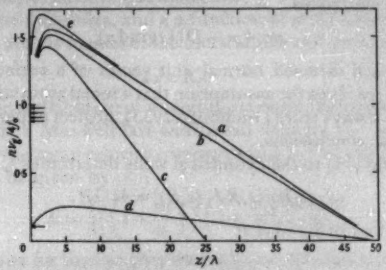
According to (10) the former is given by

$$j_e = \frac{4j\epsilon D_0}{v_M R} \frac{\sum_\nu f_\nu A_\nu J_0(A_\nu r/R) \coth(A_\nu d/R)}{\sum_\nu f_\nu J_0(A_\nu r/R)} \qquad \ldots\ldots(21)$$

and the latter by the integral of j_e over the cathode, or

$$F = \frac{4I\epsilon D_0}{v_M R} \frac{\sum_\nu f_\nu J_1(A_\nu) \coth(A_\nu d/R)}{\sum_\nu f_\nu J_1(A_\nu)/A_\nu} \qquad \ldots\ldots(22)$$

where I is the integral of j over the cathode.



Typical axial density distributions for aluminium in neon, calculated from equations (11) and (19), for the following conditions:

(*a*) $2R = 500\lambda$, $d = 50\lambda$, $v_g/v_0 = 1$.
(*b*) as for (*a*), but $2R = 250\lambda$
(*c*) as for (*a*), but $d = 25\lambda$
(*d*) as for (*a*), but $v_g = v_0/10$
(*e*) $2R = 1000\lambda_1$, $d = 100\lambda_1$, $v_g/v_0 = 1$, $\lambda_1 = \lambda/2$.

The ordinate is the non-dimensional quantity $nv_g/4j_0$. The arrows indicate the points where the curves cut the density axis; they refer from above downwards to curves *e*, *a*, *b*, *c*, *d*. For the above curves the simple relation $f = J_0(A_0 r/R)$ has been assumed.

The density of the wall deposit in the diffusion zone as a function of z will be given by (19), on account of (14). The corresponding application of (11) for the initial zone can obviously only lead to a rough prediction of the wall deposit there.

§ 3. Conclusions and Comparison with Experiment

We can draw the following conclusions from (14), (19), (21) and (22) and the figure:

(*a*) The density distribution function $n(z, r)$ of metal atoms in the discharge is a complicated function of the coordinates, which is only separable in z and r if $f(r/R)$ is predominantly determined by a single Bessel component. In the general case the distribution $n(r)$ is strongly dependent on distance from the cathode, and approaches $J_0(A_0 r/R)$ as z and d increase. Likewise the axial density distribution $n(z)$ varies with r, tending to $\sinh\{A_0(d-z)/R\}$ as $r \to R$. A pronounced density maximum appears in front of the cathode, falling rapidly to a finite value at the cathode surface, and more gradually towards the anode, where it has the value zero. The latter fall occurs almost linearly if $R \gg d$, otherwise this part of the curve is more or less concave upwards. The density maximum decreases with decreasing R and decreasing d. It decreases also with decreasing pressure and increasing mass of the metal atoms, at the same time moving farther from the cathode surface. The influence of the emission velocity v_0 is specially striking. The density everywhere is considerably diminished by increase in v_0 whilst the maximum broadens and recedes from the cathode.

(*b*) Let us call the ratio j_0/j and F/I partial and total yield factors, and denote them by κ_p and κ_T respectively. Then (21) and (22) show that both yield factors must be very small compared with unity, since $\lambda \ll R$. This means that by far the greater number of metal atoms diffuse back to the cathode. Equations (3), (4) and (20) show in addition that the yield factors are dependent markedly on the mass and emission velocity of the metal atoms. The dependence of κ_p and κ_T on anode distance d is complicated, and given by simple relations only when d is small or large. For $d \gg R/A_0$ both yield factors approach limiting values. The limiting value of κ_p is a function of r.

If A_m denotes the greatest zero value of the Bessel components which contribute significantly to (22) and (23), then if $d \ll R/A_m$, κ_p and κ_T become identical and are proportional to $1/d$. This restriction on d corresponds to the simple linear approximation to the problem, and gives a strict criterion for its validity. In all other cases κ_p and κ_T are involved functions of d, in which κ_p also depends on r. A general simplification occurs only if $f(r/R)$ is effectively identical with one of the Bessel components $J_0(A_r/R)$. The radial dependence of κ_p then disappears, κ_p and κ_T become the same, and their variation with d is described by a simple coth law.

An exact comparison between the conclusions drawn in (*a*) and (*b*) and experiment would require better knowledge of parameters such as f, and v_0. In addition we should need more data obtained from experiments in which the simple discharge conditions postulated have been approached. In no instance, therefore, can a complete quantitative test be made. It is nevertheless possible to show that there is order of magnitude or qualitative agreement between theory and experiment.

The only experimental investigation known to us which allows conclusions to be drawn about the dependence of the density distribution n on z and r is a comprehensive investigation made by Guntherschulze (1942) on sputtering of copper and silver in hydrogen. Guntherschulze's curves for the density of the wall deposits (figure 16 of his paper) show close resemblance to the curves in the figure given in this paper. The change in position and form of the maxima of Guntherschulze's curves with voltage also agrees with our results for the effect of increase in emission velocity v_0. In accordance with our theory, the anode sections of Guntherschulze's curves show a greater departure from linearity than the curves in our figure, as he has studied a tube with a small ratio of R to d. Guntherschulze's measurements of the radial variation of the density of the anode deposit cannot be interpreted fully without knowledge of $f(r/R)$. The curves (figures 4 and 5 of his paper) do, however, show the expected dying out of the higher Bessel components as d increases.

The predicted influence of v_0 is of particular significance for measurements of rate of sputtering as a function of cathode fall V. Guntherschulze (1942) has pointed out that a large discrepancy apparently exists between von Hippel's sputtering experiments in glow discharges (1926) and the corresponding low-pressure experiments by himself and Meyer (1930). Whilst, according to von Hippel, the intensity of sputtering passes through a maximum as V increases, it increases approximately linearly with V according to Guntherschulze and Meyer's measurements. Guntherschulze has attempted to account for this discrepancy by electrostatic repulsion of the discharge from the walls. We see, however, from (20) that the two sets of experiments are not necessarily in conflict. On the contrary, with the linear law for j found by Guntherschulze and Meyer a maximum might occur in von Hippel's experiment. This is because an increase in V will almost certainly lead to an increase in v_0, which will cause a decrease in ϵ, and so

gives the possibility of a maximum. It cannot, however, be stated definitely whether or where a maximum occurs without knowing v_0 as a function of V.

We know of no measurement of the partial yield factor κ_p. A measurement by Guntherschulze (1926) verifies the asymptotic value of κ_T for large d. Other experiments by Blechschmidt and Guntherschulze check with remarkable accuracy the d^{-1} law of the linear approximation for small d, although the condition $d \ll R/A_m$ was not necessarily satisfied. The order of magnitude of κ_T found experimentally in almost all cases is λ/R, as predicted.

Seeliger (1942) has obtained an absolute value of the yield factor κ_T from the glow discharge experiments of Guntherschulze (1926a) and the low-pressure experiments of Guntherschulze and Meyer (1930). He compares this experimental value with the value obtained from his simplified linear diffusion theory, and finds an order of magnitude discrepancy, the latter being 16 times smaller than the former. In order to explain this discrepancy Seeliger suggested that a substantial amount of sputtering was produced by impact of fast neutral particles. There is no doubt that neutral particles can contribute to sputtering. It seems possible, however, that the discrepancy found by Seeliger can at least be partly accounted for in another way from our theory. Adopting, like Seeliger, the linear approximation, there results between the factor $\kappa_T(S)$ calculated by Seeliger, and our value κ_T, the relation

$$\kappa_T = \frac{4}{3}\left\{\frac{D_0}{D_c} + \frac{3K}{4}\frac{\ln(D_c/D_0)}{(D_c-D_0)/D_0}\right\}\frac{(1+M/m)^{1/2}}{1+\alpha}\kappa_T(S)$$
$$= \phi\kappa_T(S), \text{ say.}$$

The discharge was one through hydrogen, with a silver cathode, and a cathode fall of 550 v. It is therefore likely that D_c was not much greater than D_0. In this case ϕ will be between about 15 and 20, according to whether the hydrogen was molecular or atomic, which is of the order of Seeliger's discrepancy.

Summarizing our conclusions, we believe we have shown that cathode sputtering in glow discharges is more complex than has been hitherto assumed. The theoretical complications appear, however, to be so great that sputtering experiments in which diffusion of the metal occurs are unlikely ever to give precise data about the atomic mechanism of cathode disintegration.

Acknowledgment

One of us (G.E.) is indebted to the Warren Committee of the Royal Society and to the Queen's University of Belfast for making his stay in Belfast possible.

References

Baum, T., 1927, *Z. Phys.*, **40**, 686.
Chaudhri, R. M., and Baqui, M. A., 1952, *Proc. Phys. Soc.* B, **65**, 324.
Childs, E. C., 1930, *Phil. Mag.*, **9**, 529.
Ditchburn, R. W., 1933, *Proc. Roy. Soc.* A, **141**, 169.
Glockler, G., and Lind, S. C., 1939, *The Electrochemistry of Gases and Other Dielectrics* (New York : John Wiley).
Guntherschulze, A., 1926, *Z. Phys.*, **36**, 563; 1926a, *Ibid.*, **38**, 575; 1942, *Ibid.*, **119**, 79.
Guntherschulze, A., and Meyer, K., 1930, *Z. Phys.*, **62**, 607.
von Hippel, A., 1926, *Ann. Phys., Lpz.*, **81**, 1043.
Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford : Clarendon Press).
Seeliger, R., 1942, *Z. Phys.*, **119**, 482.
Seeliger, R., and Sommermeyer, K., 1935, *Z. Phys.*, **93**, 692.
Townes, C. H., 1944, *Phys. Rev.*, **65**, 319.

A Rotating Interferometer for the Measurement of the Directions of Arrival of Short Radio Waves

By H. A. WHALE
Seagrove Radio Research Station, Auckland, New Zealand

Communicated by P. W. Burbidge; MS. received 15*th December* 1953, *and in amended form* 29*th March* 1954

Abstract. A method of measuring the bearing angles and the angles of elevation of incoming radio waves is described. In this method two aerials are used, one of which moves in a circular path several wavelengths in diameter. The phases of the radio signals from the two aerials are compared in a system which does not depend on their relative amplitudes and it is shown that this method enables measurements to be made even when there is considerable differential fading between the two aerials. Some experimental records are given yielding the angles of arrival of radio waves.

§ 1. Introduction

The direction of arrival of a plane radio wave may be specified by its azimuthal (bearing) angle and its angle of elevation. While considerable success has been achieved in the measurement of bearing angles under very varied conditions, the associated measurement of the angle of elevation has proved more difficult (except in the case of very short waves where radar techniques may be employed). The first measurement of the angle of elevation of an incoming wave was, of course, made in Appleton's frequency-change experiment. A method due to Pawsey (1935) in which the angle of elevation was deduced from a comparison of the signals received in two different aerial systems was the fore-runner of several similar methods. Later methods, for example that of Bramley and Ross (1951), have depended on the accurate measurement of the phase differences between signals received in pairs of aerials spaced some distance horizontally or vertically. The problem in these cases is then that of accurately measuring and recording phase differences.

When the bearing of an incoming wave is known or has been found, the angle of elevation may be deduced from a measurement of the wavelength of the wave along the ground on the known bearing. This wavelength may be found either from a measurement of the phase difference between the signals received at two points on the bearing line or from a measurement of the separation along the bearing line at which the signals received from the two aerials have zero phase difference.

While the accurate measurement of phase differences requires complex equipment, the determination of the distances apart at which the aerials must be placed for the signals to have some specified phase relationship is comparatively simple. Using this method, wavelengths along the ground may be stepped off and the angle of elevation of the incoming wave determined. In both of these methods, care must be taken to ensure that full wavelengths are not 'lost' and ambiguities thus introduced into the measurements.

§ 2. The Comparison of Two Fading Signals

When two radio signals of the same frequency and amplitude are added together, the amplitude of the resultant signal depends on the phase difference between the two signals. In particular, the amplitude of the resultant signal is zero when the phase difference is exactly 180°. If the two signals are not of the same amplitude, the resultant amplitude will never be zero, but will be a minimum when the phase difference is 180°.

It has been shown by Ratcliffe and Pawsey (1933) and confirmed by many more recent workers that the amplitudes of radio waves from the same source received via the ionosphere at two places which are some distance apart usually exhibit a certain amount of independent fading.

In such cases the amplitude of the signal resulting from the addition of the signals from the two aerials in the out-of-phase condition will seldom be zero. The degree of difference between the amplitudes of the signals received in the two aerials is commonly specified in terms of the correlation coefficient R, which is equal to 1 when the amplitudes of the two signals are always the same and is equal to zero when the amplitudes are completely uncorrelated, i.e. when the two signals show independent random fluctuations of amplitude.

Measurements have been made by many workers of the way in which the coefficient R varies as the spacing between the two aerials is changed. From their experimental values may be deduced the sharpness of the minima to be expected in a hypothetical experiment in which two receiving aerials are moved apart along a bearing line and the signals from each added together in an attempt to measure the wavelength along the ground. The fading of the radio waves may be thought of as being due to the simultaneous reception of an aggregate of waves with directions clustering around a mean direction. In such a case the fluctuations in the amplitude difference between the signals received in the two aerials will be accompanied by fluctuations in the phase difference between the two signals. Bramley (1951) has shown that $|\overline{\phi}|$, the mean phase deviation from the mean phase difference, is given by the relation

$$\cos|\overline{\phi}| = R. \qquad \ldots\ldots(1)$$

A reasonable value of R for two aerials about fifty metres apart has been found to be about 0·8. In this case the mean phase deviation will be about 35°. This phase deviation implies a residual mean signal when the two signals are added in the nominally out-of-phase condition. If the mean amplitude of each of the two signals is $\bar{A}$, this residual signal will give a contribution to the mean value of the resultant amplitude of $\bar{A}\sin|\overline{\phi}|$, i.e. about $0{\cdot}6\bar{A}$.

Similarly, the mean value of the resultant amplitude in the in-phase condition will be reduced slightly, in this case from $2\bar{A}$ to about $1{\cdot}9\bar{A}$.

Assuming $R = 0{\cdot}8$, typical forms of the probability distributions $p_{\text{amplitude}}$ of the amplitudes of (i) the assumed individual input signals, (ii) the resultant signal in the nominally in-phase condition, and (iii) the resultant signal in the nominally out-of-phase condition are given in figure 1 (*a*). For $R > 0{\cdot}8$, the spreads will be less than shown, while for $R < 0{\cdot}8$, the distributions will be broader than those shown. These curves are derived from a set of curves given by Rice (1945).

From these probability distribution curves it can be seen that the minima indicating the out-of-phase positions for a pair of spaced aerials will be very indefinite in the presence of some degree of independent fading between the two

aerials. As an example, the curve of figure 1 (*b*) depicts the type of resultant amplitude curve which might be expected in our hypothetical experiment in which the two aerials are moved apart along a bearing line and the two signals added together. The horizontal scale ϕ is in degrees of mean phase difference between the two aerials and the correlation coefficient R has been assumed to remain constant at about 0·8 even though the spacing between the two aerials changes during the course of the experiment. The 'average' curve is the curve which would be obtained by taking the mean of a very large number of such experimental records, the approximate spread of each record being enclosed by the broken lines.

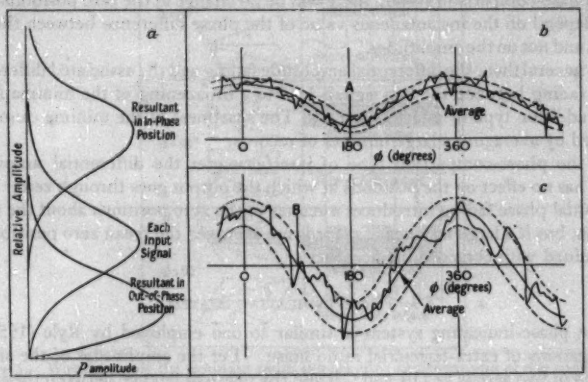


Figure 1.

(*a*) Approximate amplitude distributions of two fading radio signals when measured singly, when added in the nominally in-phase condition and when added in the nominally out-of-phase condition.

(*b*) Approximate interference pattern produced when the two fading signals are added together while the phase difference between them is changed steadily.

(*c*) Approximate interference pattern produced when the two fading signals are combined in the phase-comparison interferometer.

In the absence of differential fading between the two aerials a zero would be obtained at A, where the phase difference between the signals from the two aerials is 180°. However, in the presence of differential fading between the two aerials this zero is blurred, even on averaged records, the blurring being mainly due to the residual signal in the resultant arising from the 35° phase deviation. However, it is possible to measure the phase difference between a pair of signals by methods other than simple addition of the signals. In the system to be described we obtain an output voltage which is approximately proportional to $A_1A_2 \cos \phi$ where A_1 and A_2 are the amplitudes of the two signals and ϕ is the phase difference between them. The important advantage of this system over the previous one is that, irrespective of the values of A_1 and A_2, the output voltage goes through zero when $\phi = \pm 90°$. The differential phase fading of mean value 35° will introduce a mean scatter (of about 35° each side of the 90° positions) in the positions of the zero points obtained in each experiment in which the two aerials are moved away from each other along

a bearing line and the output voltage recorded. A record obtained in one such experiment might then look like the one drawn in figure 1 (*c*). If a large number of such experimental records are obtained and the results averaged, the mean curve will be the curve labelled 'average' in this figure. This curve passes through zero at exactly $\phi = 90°$, 270° etc.

As well as the differential fading, the extent of which is specified by the coefficient R, there will be in most cases considerable coherent fading at the two aerials, i.e. both signals will rise and fall together. This fading will, in general, have the effect of broadening the distributions in figure 1 (*a*) so that interference pattern in the experiment illustrated in figure 1 (*b*) will become less definite. However, in the phase-comparison system there will be no change in the zero positions since these depend on the instantaneous value of the phase difference between the two signals and not on the amplitudes.

In general then, the differential amplitude fading and the associated differential phase fading between the two signals lead to a broadening of the minima in the signal addition type of interferometer. The sharpness of the minima cannot be regained by averaging a large number of records.

In the phase-comparison type of interferometer the differential amplitude fading has no effect on the positions at which the output goes through zero. The differential phase fading introduces a scatter of the zero positions about the mean position, but if a large number of records are averaged the mean zero position can be obtained with considerable accuracy.

§ 3. Phase-Indicating System

The phase-indicating system is similar to one employed by Ryle (1952) in investigations of extra-terrestrial radio noise. Let the amplitudes of the signals received in two aerials be OA and OB and the phase difference between them be ϕ as in figure 2. If these two signals are added, the resultant amplitude is r_1 and if

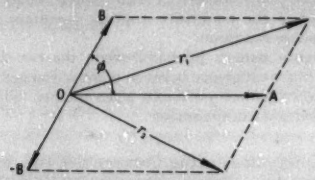


Figure 2. Basis of the switching system for measuring the phase difference between two radio signals.

they are subtracted, i.e. one of them is reversed in phase, the resultant amplitude is r_2. It is easily shown that r_1 is greater than r_2 when the magnitude of the phase difference is less than 90° and that r_1 is less than r_2 when the magnitude of the phase difference lies between 90° and 180°.

Successive addition and subtraction of the two signals may be carried out at a low switching frequency (conveniently at about 50 c/s), causing a modulation to appear on the resultant output signal as drawn in figure 3 (*b*). This resultant signal is amplified in a radio receiver and rectified in the conventional way to give an audio output at the switching frequency. After passing through a narrow-band filter, the audio output signal will be a sine-wave at the switching frequency.

The forms of the modulated signal and the filtered audio output signals for the cases $\phi<90°$, $\phi=90°$ and $\phi>90°$ are shown in figure 3. The reference waveform and the audio output are in phase when $\phi<90°$ and out of phase when $\phi>90°$. In figure 3 (*a*) is shown the reference waveform at the switching frequency. This is obtained from a small a.c. generator connected directly to the shaft of the

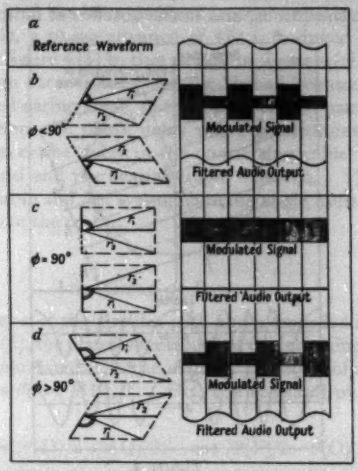


Figure 3. Waveforms of the resultant radio signals after switching and of the audio output signals as the phase difference between the two input signals from the aerials changes.

switching mechanism. If the reference waveform and the audio output are compared in a conventional phase-sensitive-rectifier circuit a voltage will be obtained whose sign reverses when the phase difference between the two low-frequency waveforms changes from 180° to 0° or vice versa, i.e. when the phase difference between the two radio signals from the aerials changes from less than 90° to greater than 90° or vice versa. It is this voltage which is obtained from the phase-sensitive-rectifier circuit which is recorded to give an interference pattern similar to the one depicted in figure 1 (*c*).

§ 4. Rotating Phase-Comparison Interferometer

In the Introduction it was assumed that the bearing angle of the incoming signal was known and it was stated that the angle of elevation could be found from a measurement of the wavelength along the ground along the known bearing. In practice, both the bearing angle and the angle of elevation are unknown so that both must be measured. In the present interferometer two receiving aerials are used. One of these is fixed and the other is allowed to move around a horizontal circle several wavelengths in diameter. For convenience, the fixed aerial is placed at the centre of this circle, but this central position is not essential as the signal from it is used only as a reference.

As an example, consider the plan view of such a system shown in figure 4 (*a*). The fixed aerial is at the centre, A, and the moving aerial follows the circular path BCD . . . QRB. In this example the diameter of the circular path is chosen, for simplicity, to be a whole number of times the wavelength of the incoming wave resolved along the ground, λ_g. It is also assumed, for convenience, that the radio signals from the aerial at B can be transmitted to A with a whole number of cycles delay. Under these conditions, and in the absence of fading, the output voltage

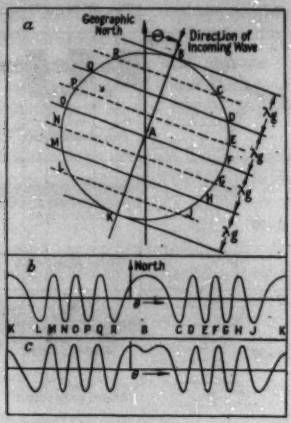


Figure 4.

(*a*) Plan view of the rotating interferometer.
(*b*) Interference pattern produced when there is no delay in the cable connecting B to A.
(*c*) Interference pattern produced when a 45° phase advance is introduced into the cable connecting B to A.

from the phase-comparison system as the aerial B moves along the path BCDE . . . QRB will have the form shown in figure 4 (*b*). The way in which the bearing angle and the angle of elevation of the incoming radio wave can be deduced from the positions of the maxima and minima of this pattern and the effects of a change in the delay incurred in transmitting the signal received at B to the comparison system at A are discussed in §§ 5 and 6.

§ 5. Bearing Angle

The record shown in figure 4 (*b*) has four centres of symmetry, viz. B, F, K and O. Two of these are obviously places where the rate of change of phase difference with the angular displacement θ of the aerial is large, i.e. at F and O the maxima and minima are closely spaced. It can be seen that the line BK in figure 4 (*a*) passes through the 'broad' centres of symmetry where the rate of change of phase difference with the angle θ goes through zero and defines the bearing angle Θ measured from a reference direction, say, geographic north.

The sense of the bearing is still undefined, i.e. the record as it stands does not contain the information whether the direction of the signal is BK or KB. It was assumed, for convenience, that there was a whole number of cycles delay in the

link connecting the aerial at B, C, D . . . to the phase comparison equipment at A. Any reduction in the delay in the transmission line connecting B to A will result in a shifting of the lines QD, OF, MH, etc. in figure 4 (*a*) away from the direction from which the radio wave is arriving since the phase advance introduced in the cable must be compensated by a phase delay in the signal received in the moving aerial at B, C, D, . . ., etc. This shift of the pattern may be used for determining the sense of the bearing which has been deduced from the symmetries of the recorded pattern. For example, a phase advance of 45° introduced into the link would change the pattern in figure 4 (*b*) to that drawn in figure 4 (*c*).

It will be seen from these figures that the 'broad' centre of symmetry of the pattern still defines the bearing angle, so that this modified pattern is as useful as the one in figure 4 (*b*) where the phase delay was a whole number of cycles. It is easily seen that whatever the delay in the connecting cable, the broad centre of symmetry can be found and the bearing angle deduced. The only condition is that the delay in the cable and the average bearing angle must remain constant for the time it takes to make the record.

§ 6. Angle of Elevation

Given the diameter d of the circular path and a recorded interference pattern, say the one in figure 4 (*b*), the phase wavelength of the incoming wave resolved along the ground may be calculated. If the angular positions of the points D, F, H; M, O, Q are given as $\theta(\mathrm{D})$, $\theta(\mathrm{F})$, . . ., etc., the required phase wavelength λ_g is seen to be given by

$$\lambda_g = \tfrac{1}{2}d[\cos\{\theta(\mathrm{D})/2-\theta(\mathrm{Q})/2\}-\cos\{\theta(\mathrm{F})/2-\theta(\mathrm{O})/2\}]$$
$$= \tfrac{1}{2}d[\cos\{\theta(\mathrm{F})/2-\theta(\mathrm{O})/2\}-\cos\{\theta(\mathrm{H})/2-\theta(\mathrm{M})/2\}]. \quad \ldots\ldots(2)$$

In an actual case, the average of these values would be taken.

The wavelength along the ground depends only on the angle of elevation of the incoming wave and not on the electrical properties of the ground. For, if δ is the angle of elevation of an incoming wave, the field at an aerial will be the resultant of the direct wave and the wave reflected from the ground (the reflected wave having an angle of elevation $-\delta$).

The direct and reflected waves will have wave functions

$$A\exp\{k(x\cos\delta - y\sin\delta)\}$$

and

$$rA\exp\{k(x\cos\delta + y\sin\delta)\}$$

where r = ground reflection coefficient (complex), A = amplitude of the direct wave, x = distance measured horizontally in the plane of incidence, y = distance measured vertically upwards, $k = 2\pi/\lambda$ and λ = free space wavelength of the direct wave.

Let y be measured from the horizontal plane in which the aerial moves so that, at the aerial ($y=0$), the resultant field is

$$(1+r)A\exp\{kx\cos\delta\}.$$

From this expression it can be seen that λ_g, the wavelength in the x-direction, is given by

$$\lambda_g = \lambda/\cos\delta, \quad \ldots\ldots(3)$$

so that the angle δ may be found from a measurement of the quantity λ_g.

§ 7. Recording Methods

The pattern in figure 4 (*b*) was idealized in that it was assumed that there was no fading of the incoming radio wave. When there is fading, individual patterns become distorted and it is necessary to take the average of many such patterns in order to find the mean direction of arrival accurately.

The averaged pattern is easily obtained if individual patterns are recorded one above the other on a circular drum which rotates with the moving aerial. It is convenient to use variable density records in which the density of the pattern is proportional to the voltage output from the phase comparison equipment. The variable density pattern is produced by a dotting pen which makes dots on the rotating drum at a rate proportional to the output voltage from the equipment.

Three sections of record are reproduced in figure 5. That in figure 5 (*a*) was made from a simple signal-addition interferometer, i.e. a system which yields a record of the type which has been drawn in figure 1 (*b*). The others, in figures 5 (*b*) and 5 (*c*), are from the phase-comparison interferometer in which the records are of the form drawn in figure 1 (*c*). All 'negative' dotting rates have been suppressed, i.e. the pen makes dots only when the output voltage is in one direction. The increased clarity of the phase-comparison system is obvious from the records.

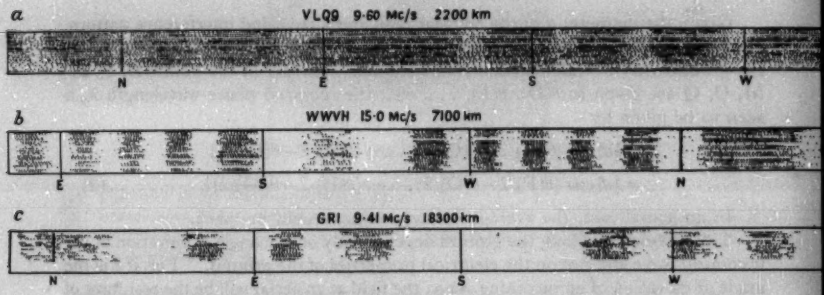


Figure 5.

a. Experimental interference pattern obtained with a signal-addition type of rotating interferometer. (*Note*. The frequency in this observation was 9·66 Mc/s, not 9·60 Mc/s.) *b* and *c*. Experimental interference patterns obtained with the phase-comparison type of rotating interferometer.

In these measurements, the aerial spacing (i.e. the radius of the circular track) was sixty metres. With this spacing the correlation coefficient R is usually about 0·8 and it has been found possible to make measurements of the directions of arrival of radio waves from most parts of the world, including the B.B.C. transmissions from the United Kingdom. Although the records will be discussed further in another paper, it is interesting to note here that so far they often appear surprisingly simple, indicating that the waves are clustered about a single main direction.

§ 8. Equipment

The basis of the system has been described in the preceding sections. A block diagram of the complete equipment is given in figure 6 (*a*). Both fixed and moving aerials are horizontal loop aerials about 16 feet above the ground. In this way

all-round reception of the horizontally polarized component of the incoming wave is obtained. The loop aerials, which are untuned, feed into wideband amplifiers covering the range which is being used at the moment, viz. 5–25 Mc/s.

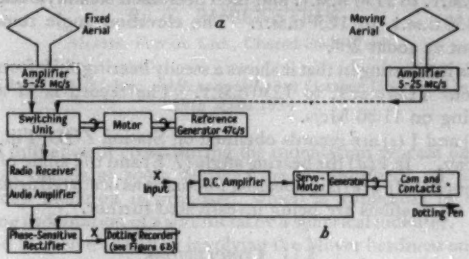


Figure 6.

a. Block diagram of equipment.
b. Essentials of the dotting pen recorder.

The switching unit contains a motor-driven reversing transformer and a mixing transformer which carry out the additions and subtractions indicated in figure 2 at about 47 c/s. The output from the phase-sensitive rectifier is used to control the speed of a d.c. motor generator as indicated in figure 6(*b*). The impulses for the magnetically operated dotting pen are obtained from a cam and contacts mounted on the shaft of the combined motor generator while a clutch-operated pair of contacts ensures that the pen operates when the motor is running in one direction and not in the other.

§ 9. Mechanical Arrangements

The fixed horizontal loop aerial is mounted on a central equipment hut. Through the centre of this loop extends a vertical rotatable mast which supports the power cables, high-frequency cable and the steering wires to the moving aerial. Power from the mains supply is fed to these cables and the signal from the moving aerial is picked off through a system of slip-rings at the base of the rotatable mast.

The moving horizontal loop aerial is mounted on a three-wheeled carriage. The carriage is driven by an electric motor at speeds up to about ten miles per hour and is maintained on its circular path by steering wires attached to a steering arm in front of the single front wheel. This wheel is pivoted and is sprung so that the carriage tends to travel in a spiral of increasing radius. When the steering wire, which is firmly attached to the rotatable mast on the central hut, tightens, the carriage can no longer spiral outwards and then maintains a circular path.

§ 10. Sample Records and Results.

Five sample sections of record are reproduced in Plate I. The records have been unrolled from the cylindrical recording drum so that each covers (horizontally) a range of 360°. The timing marks can be seen on the records, time increasing vertically upwards as indicated.

Plate I (*a*) represents measurements made on the signal received from Station WWVH, Hawaii, on 15·00 Mc/s. The bearing angle and angle of elevation of this signal have remained fairly steady at 29° E and 18° respectively during the period of recording.

Plate I (*b*) has been chosen to illustrate the variations in bearing angle which sometimes occur. The signal was from a station on 9·41 Mc/s (probably GRI) and shows how the bearing angle increased steadily from 8° E to 30° E during the period 1030 G.M.T. to 1130 G.M.T. and then decreased steadily back to 8° E during the period 1130 G.M.T. to 1230 G.M.T. The elevation angle remained approximately constant at about 20°.

Plate I (*c*) is interesting in that it shows a steady bearing shift from 62° E to 48° E during the hour 1130 G.M.T. to 1230 G.M.T. The record is of an unidentified station operating on 11·80 Mc/s.

Plates I (*d*) and I (*e*) are records obtained on Station ZQD, Fiji, taken over the same time of day. In I (*d*) the bearing angle (7°E) and the angle of elevation (15°) did not alter much. In I (*e*) large bearing fluctuations occurred around 1300 G.M.T. Such bearing fluctuations are being investigated further.

§ 11. Conclusions

A method of phase-comparison interferometry suitable for the measurement of the direction of arrival of radio waves which have been reflected from the ionosphere has been described. The method enables accurate average values of bearing angle and angle of elevation to be obtained in the presence of considerable differential fading between the two aerials.

Acknowledgments

The rotating radio interferometer was built by the staff and students working on the Seagrove Radio Research Station. This Station is operated by the Department of Physics, Auckland University College, and receives a grant from the N.Z. Department of Scientific and Industrial Research.

References

Bramley, E. N., 1951, *Proc. Instn Elect. Engrs*, **98**, Pt. III, 22.
Bramley, E. N., and Ross, W., 1951, *Proc. Roy. Soc.* A, **207**, 251.
Pawsey, J. L., 1935, *Proc. Camb. Phil. Soc.*, **31**, 125.
Ratcliffe, J. A., and Pawsey, J. L., 1933, *Proc. Camb. Phil. Soc.*, **29**, 301.
Rice, S. O., 1945, *Bell Syst. Tech. J.*, **24**, 102.
Ryle, M., 1952, *Proc. Roy. Soc.* A, **211**, 351.

The Influence of Creep on the Measured Hardness of Soft Metals

By C. RUBENSTEIN
Messrs. Ferodo Ltd., Chapel-en-le-Frith

Communicated by R. C. Parker; MS. received 13th October 1953, and in amended form 29th April 1954

Abstract. The influence of creep during the indentation process on the measured hardness of soft metals at room temperature is described. A modification to Dushman's creep equation is proposed, which is applicable to the conditions obtaining during indentation of a soft metal by a spherical indenter. The modified equation is used to derive a relation involving the Meyer hardness and the diameter of the indenter, which agrees satisfactorily both with published experimental data and with data obtained by the author. Numerical values of the Dushman constant α for lead and indium, obtained from this analysis of hardness measurements, are in reasonable agreement with values derived in other ways.

§ 1. Introduction

Hatch and Whittaker (1950) have shown that the value of the Meyer hardness obtained when a steel ball of given radius is pressed into an indium or lead flat is the same as that obtained when a hemispherical specimen of lead or indium of the same radius is pressed against a hard flat surface. This interchangeability of roles permits the application of the results of Parker and Hatch (1950) to the consideration of the factors influencing the measured hardness of lead and indium. These authors measured the area of contact between hemispherical specimens of lead and indium loaded against a glass plate under various loads, each load being applied for the same time, and found that the area of contact varied directly with applied load, i.e. the Meyer hardness was independent of load. The value of the Meyer hardness was, however, dependent upon the diameter of the ball, being greater for smaller diameter balls.

It has been pointed out by Hatch and Whittaker (1950) that the Meyer–O'Neill equation

$$L = \frac{K d^n}{D^{n-2}} \qquad \ldots\ldots(1)$$

where L is the applied load, d the indentation diameter, D the ball diameter and K, n are constants for a given pair of materials, requires that when $n=2$ (which obtains with lead and indium), the Meyer hardness ($4L/\pi d^2$) should be independent of ball diameter, so that equation (1) is not supported by the experimental work of Parker and Hatch (1950).

In view of the occurrence of creep when soft metals are stressed at room temperature, it is not surprising that Meyer's law is inapplicable to lead and indium, for O'Neill (1934) states that " the Meyer analysis by the ball test is only properly applicable under conditions where viscous flow is not very pronounced, as the ball should reach static equilibrium under the load ".

Previous work, detailed above, has indicated the occurrence of creep during the hardness testing of soft materials. In the following, an attempt is made to determine quantitatively the extent to which the measured hardness is dependent on the creep occurring during indentation.

§ 2. Theoretical Considerations

Dushman *et al.* (1944) and Chalmers (1936, 1937) have shown that for small deformations at very low stresses the creep is of a viscous type, i.e. the creep rate is constant for a constant applied stress. The strain rates in these experiments were of the order 10^{-8}–$10^{-6}\,\mathrm{sec}^{-1}$ and any cross-sectional changes were therefore negligible, the applied stress being taken to be constant.

Dushman *et al.* proposed on the basis of theoretical and experimental results the equation:

$$\frac{d\epsilon}{dt} = Be^{\alpha p} \qquad \ldots\ldots(2)$$

where ϵ is the creep strain, p the applied stress and α, B are temperature dependent constants, B having the dimensions $(\text{time})^{-1}$.

For cylindrical specimens, both the stress and the strain are defined independently and are related only by the experimentally observed fact that, to a good approximation, there is no permanent change in volume during plastic deformation. In indentation hardness measurements using a ball indenter, the Meyer hardness can be taken as a representative value of the stress, while the strain can be represented by a function of d/D (where d is the diameter of the indentation and D is the diameter of the indenter) which is dimensionless, and completely specifies the shape of the indentation at any stage of the deformation. Thus both the stress and the strain are defined as functions of the diameter of the indentation and are consequently inter-related. Because of this inter-relation of stress and strain, a creep equation applicable to the results obtained with cylindrical specimens must undergo some modification before it can be applied to the creep processes occurring during indentation. Considerations leading to a suitably modified form of Dushman's equation are discussed in the following.

By analogy with the definition of strain in compression, the strain during indentation can be taken as A/A_0 where A is the projected area of the indentation and A_0 $(=\frac{1}{4}\pi D^2)$ is the diametral area of the spherical indenter. Equation (2) becomes

$$\frac{d\epsilon}{dt} = \frac{d}{dt}\frac{A}{A_0} = \frac{1}{A_0}\frac{dA}{dt} = -\frac{L}{p^2 A_0}\frac{dp}{dt} \qquad \ldots\ldots(3)$$

since $p = L/A$.

Moore and Tabor (1952) determined the rate of decrease of Meyer hardness with time in indium and obtained, empirically, an expression for $dp/dt = -1/\alpha t$ where t is the time for which the load was applied and α is the constant used in Dushman's equation.

Substituting for dp/dt in equation (3),

$$\frac{d\epsilon}{dt} = \frac{\mathrm{L}}{p^2 A_0 \alpha t}. \qquad \ldots\ldots(4)$$

The experimental work referred to above has shown that in the case of indentation of lead and indium by a ball using loads above 150 g the stress is independent of the load. The left-hand side of equation (4) is proportional to

the applied load and in order to make the right-hand side of this equation proportional to the load, B must be replaced by CL where C is a constant having the dimensions (time)$^{-1}$. Thus considering the interdependence of the stress and the strain and the experimentally observed relation between the applied load and the projected area of the indentation, Dushman's equation must be modified to

$$d\epsilon/dt = CLe^{\alpha p}$$

when applied to the creep occurring during indentation by a spherical indenter. Substituting for $d\epsilon/dt$ from equation (4)

$$\frac{\mathrm{L}}{p^2 A_0 \alpha t} = CLe^{\alpha p}$$

whence

$$e^{\alpha p} = \frac{1}{Gp^2 D^2 \alpha t}$$

where $G = \frac{1}{4}\pi C$, i.e.

$$p = \frac{1}{\alpha} \ln \frac{1}{Gp^2 D^2 \alpha} - \frac{1}{\alpha} \ln t.$$

For a constant time of loading, this leads to the relation

$$p = \frac{1}{\alpha} \ln \frac{1}{Gp^2 D^2 \alpha} - \text{constant.} \qquad \ldots\ldots (5)$$

To determine the extent to which equation (5) is obeyed, experiments on the variation of Meyer hardness with indenter diameter using indium, zinc and tin specimens were made. These experiments are described in the following section.

§ 3. Experimental

As has been shown in the previous section, equation (5) can only be applied (*a*) when the Meyer index $n = 2$, and (*b*) when creep occurs solely by a viscous-type mechanism. Condition (*a*) is fulfilled by all the metals examined in this paper under the particular test conditions imposed. As for condition (*b*), it is known that a viscous-type mechanism can occur not only in the micro-creep range but also at higher strain-rates, if the temperature is sufficiently high. This mechanism corresponds to the *K*-flow of Andrade (1914), and appears to occur predominantly by movement at the grain boundaries. A creep component of this nature would therefore be expected under the conditions of the present experiments, even though the highly localized stressing system is distinctly different from that imposed in conventional creep testing.

The final justification for the assumption that condition (*b*) is applicable to indentation creep must clearly rest on the agreement between theory and experiment.

Thus provided the metals tested fulfil the above conditions, equation (5) suggests that for a series of indentations made by a number of ball indenters of different diameters, the load being applied for the same time in each case, a graph of p plotted against $\log pD$ (where p is the Meyer hardness and D is the indenter diameter) should be a straight line.

3.1. *The Soft Metals Lead and Indium*

In order to verify this linear relationship, a series of indentations was made in a flat specimen of indium using ball-ended indenters fitted to a Vickers pyramid hardness machine, the indenter diameters ranging from 1 mm to $1\frac{1}{16}$ inch.

Each indentation was made in a fresh part of the surface, the load being applied for the same period of time in every case. The diameter of the indentation was measured in two mutually perpendicular directions using the microscope integral with the machine and a micrometer eyepiece, and the product of these values was taken as $\overline{d}^2$.

The results obtained for indium are shown in figure 1, in which $\frac{1}{4}\pi \times$ Meyer hardness (p') is plotted against $\log_{10} p'D$, and are in reasonable agreement with the relationship required by equation (5).

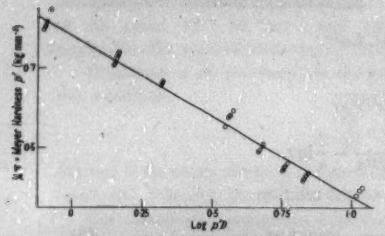


Figure 1. Variation of $\frac{1}{4}\pi \times$ Meyer hardness with the logarithm of the product of $\frac{1}{4}\pi \times$ Meyer hardness and the ball diameter for indium.

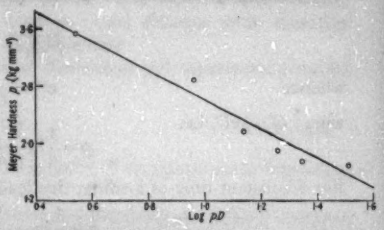


Figure 2. Variation of the Meyer hardness with the logarithm of the product of the Meyer hardness and the ball diameter for lead.

Values of p and $\log pD$ for lead, derived from the experimental results of Parker and Hatch (1950), are illustrated graphically in figure 2. These results, which showed an anomalous dependence on ball diameter when subjected to the Meyer analysis, are again in reasonable agreement with equation (5).

3.2. *The Harder Metals Zinc and Tin*

To determine experimentally the relationship between the load and the projected area of the indentation, the same method as was used for indium was applied to specimens of zinc and tin. The specimens were prepared by casting

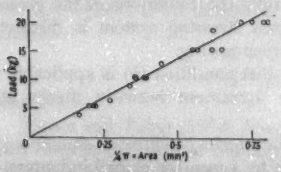


Figure 3. Dependence of $\frac{1}{4}\pi \times$ projected area of indentation on the applied load for worked zinc.

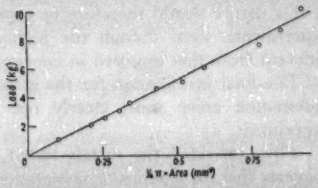


Figure 4. Dependence of $\frac{1}{4}\pi \times$ projected area of indentation on the applied load for worked tin.

from the melt, heavily working in compression and hand polishing by conventional methods. Indentations were made using a series of loads, each load being applied for the same time while the diameter of the indenter remained unaltered. The results obtained (shown graphically in figures 3 and 4) indicate a linear relationship between the applied load and the projected area of the indentation, i.e. the Meyer index $n = 2$.

Experimental verification of the relationship existing between p and $\log pD$ for these materials was obtained with a series of ball indenters of different diameters using similar methods. The results obtained are shown in figures 5 and 6 where the linear relationship between p' and $\log p'D$ can be observed.

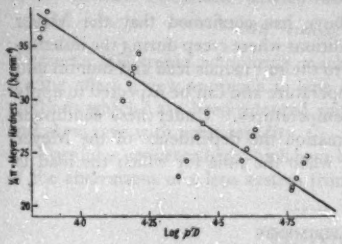


Figure 5. Variation of $\frac{1}{4}\pi \times$ Meyer hardness with the logarithm of the product of $\frac{1}{4}\pi \times$ Meyer hardness and the ball diameter for worked zinc.

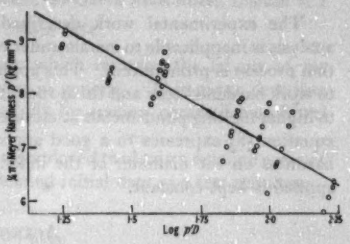


Figure 6. Variation of $\frac{1}{4}\pi \times$ Meyer hardness with the logarithm of the product of $\frac{1}{4}\pi \times$ Meyer hardness and the ball diameter for worked tin.

§ 4. Discussion of Results

In addition to predicting a linear relation between p' and $\log p'D$ equation (5) suggests that the slope of the line should be $(-4/\pi)(2 \times 2{\cdot}3026/\alpha)$, from which a value for α can be derived.

The value of α for indium derived from the line in figure 1 is $14{\cdot}6\,\mathrm{mm^2\,kg^{-1}}$. By repeated indentation experiments on indium (Rubenstein and Spurr, unpublished), a value has been obtained for α of $(14{\cdot}9 \pm 2{\cdot}3)\,\mathrm{mm^2\,kg^{-1}}$. These values show reasonable agreement with the value of $14{\cdot}0\,\mathrm{mm^2\,kg^{-1}}$ estimated from the experimental results obtained by Moore and Tabor (1952).

The value of α for lead deduced from the slope of the line in figure 2 is $2{\cdot}26\,\mathrm{mm^2\,kg^{-1}}$ which is to be compared with a value of $2{\cdot}38\,\mathrm{mm^2\,kg^{-1}}$ deduced from the creep data published by McKeown (1937), obtained using lead in the form of extruded rod.

The values of the hardness of indium, zinc and tin, reported above, were determined at room temperature. These working conditions can be compared to those of creep experiments described in the literature, in terms of a working temperature parameter θ defined as the ratio of the working temperature in degrees absolute to the melting point of the material in degrees absolute. In the following table the values of θ have been calculated for several creep determinations and these are compared with the values of θ for the experimental work described above.

Material	Range of θ (%)	Source	Material	Range of θ (%)	Source
Aluminium	59–72	D	Lead	48	PH, M
Platinum	58–76	D	Indium	45	MT, A
Constantan	45–53	D	Zinc	42	A
Ni–Mo alloys	65–74	D	Tin	57	A

D, Dushman *et al.*; PH, Parker and Hatch; MT, Moore and Tabor; M, McKeown; A, Author.

From this table it can be seen that the hardness determinations have been made at temperature conditions comparable with those used by Dushman and co-workers, within which range the creep behaviour of the metals used was shown to satisfy a viscous-type law.

§ 5. Conclusions

The experimental work described above has confirmed that the Meyer analysis is inapplicable to metals under conditions where creep during the indentation process is pronounced. This applies to the soft metals lead and indium and to work hardened zinc and tin at room temperature and can be expected to apply to higher melting point metals at elevated temperatures. Under these conditions equation (5) expresses to a good approximation the dependence of the Meyer hardness on the diameter of the indenter when the time for which the load is applied is kept constant.

Acknowledgments

The author wishes to express his thanks to members of the Ferodo Technical Division for criticisms and suggestions, especially to Dr. R. T. Spurr for initial guidance on research programme and to Mr. A. Jenkins for help with the presentation. Thanks are also due to Dr. A. J. Kennedy of the Royal Institution for helpful criticism and to Dr. R. C. Parker, Technical Director of Messrs. Ferodo Ltd., for permission to publish this paper.

References

Andrade, E. N. da C., 1914, *Proc. Roy. Soc.* A, **90,** 329.
Chalmers, B., 1936, *Proc. Roy. Soc.* A, **156,** 427; 1937, *J. Inst. Met.*, **61,** 103.
Dushman, S., Dunbar, L. W., and Huthsteiner, H., 1944, *J. Appl. Phys*, **15,** 108.
Hatch, D., and Whittaker, E. J. W., 1950, *Nature, Lond.*, **165,** 646.
McKeown, J., 1937, *J. Inst. Met.*, **60,** 201.
Moore, A. C., and Tabor, D., 1952, *Brit. J. Appl. Phys*, **3,** 299.
O'Neill, H., 1934, *The Hardness of Metals and its Measurement* (London: Chapman and Hall).
Parker, R. C., and Hatch, D., 1950, *Proc. Phys. Soc.* B, **63,** 185.

Ray Tracing on the Manchester University Electronic Computing Machine †

By G. BLACK

British Scientific Instrument Research Association, 20 Queen Anne Street, London W.1

MS. received 9th March 1954

Abstract. Electronic computing machinery which removes the labour of ray tracing offers a new approach to lens design in general. This paper describes methods which have been adopted successfully for tracing paraxial, meridional and skew rays with great rapidity.

Formulae are described which have been employed to obtain a detailed account of the aberrations of a lens system from defined initial data in a few minutes.

§ 1. Introduction

A description of the Manchester Mk2 electronic computing machine appears in a recent paper by Kilburn, Toothill, Edwards and Pollard (1953). A detailed account of the machine would be out of place in this paper but a brief survey of its construction is as follows: Two types of storage are used, electronic and magnetic. The electronic storage consists of eight cathode-ray tubes, referred to as ' pages ', each of which contains 64 ' short lines ' of 20 binary digits. Short lines can represent either numbers or instructions.

Computing proper is carried out in the 80-digit accumulator and the 40-digit multiplicand register. Addition and subtraction require 1·20 msec and multiplication 2·16 msec. The magnetic storage is exactly 64 times that of the electronic storage and takes the form of a revolving magnetic recording drum the surface of which is divided into 256 ' tracks ', each of which is equivalent to two electronic storage tubes. One or two pages can be written on to, or read from, a track by instructions requiring 94 msec for writing and 38 msec for reading or checking.

The Manchester machine has an auxiliary feature which is designed to take advantage of the repetitive nature of many calculations. These are the ' B ' tubes, which are particularly useful in optical calculations for (*a*) counting the number of times a particular operation is performed and (*b*) modifying addresses of incoming instructions, so making it possible to perform identical operations on different numbers in diverse storage locations. There are also facilities for testing the sign of numbers in the accumulator or any of the B lines, the course of subsequent calculation being determined by the positive or negative nature of the number tested.

§ 2. Input and Output

It has been found convenient in optical calculations to have a main instruction tape which supplies the machine with an account of the arithmetical and other instructions required in the calculation. These sets of instructions (routines) are read directly into the electronic storage and transferred immediately into pre-assigned magnetic storage locations.

† Extra-mural research carried out at the Computing Machine Laboratory, University of Manchester.

Information is fed into the machine by data punched on teleprinter tape which is read by a photoelectric unit at a speed of 200 characters per second. Numerical data (i.e. system constants, ray starting values, etc.) are arranged on a separate tape, and use is made of a binary–decimal conversion routine which is permanently kept in the magnetic storage. Data referring to a particular surface are read into the electronic storage after conversion and are transferred in a block to a pre-assigned position in the magnetic storage. This makes it possible to read one set of surface data at any time during calculation. All calculation is controlled by a 'master' routine which is responsible for the calling in of particular routines, and is the routine to which reference is made on the completion of any separate routine. Various routines are permanently stored in the machine for binary–decimal output, and these are usually arranged to either print results in decimal form or punch results on tape. By the punched tape method the machine can produce 15 characters per second, whereas results printed in decimal form by a teleprinter unit are produced at a speed of six characters per second. Results obtained on punched tape can of course be printed by a separate unit away from the machine.

§ 3. Optical Calculation Routines

3.1. *Paraxial Rays*

Any formulae describing the passage of a ray must remain finite and determinate in all cases, and must give rise to quantities within a certain range for flat surfaces or for surfaces of large curvature. For all paraxial rays the following formulae have been found adequate: if h is the semi-aperture of the incident pencil, u the angle of inclination of the pencil with the optical axis, c the surface curvature, and n the refractive index, then

$\Delta(hc)=\Delta(u+i)=0$, $\Delta(ni)=0$ on refraction, and $h_2=h_1-du_2$ between surfaces.

The calculation begins by reading down the data of the first surface from magnetic to electronic storage. The refraction equations are then evaluated and the transfer operation completed. At this point the input quantities for the first surface are replaced by those for the second surface obtained from the calculation. The first set of surface data is then replaced by the second using a magnetic transfer instruction, and an exactly similar calculation is performed, and so on through the system. One of the functions of the B tubes is employed in this connection. If a system has n surfaces, the number $n-1$ is placed in one of the tubes, and this number is reduced by unity every time a set of data is read down to the electronic storage. After any surface is traversed the sign of the number in the particular B line is tested and is found to be positive throughout the system until the data of the nth surface is read. On testing the sign after calculation through this surface it is found to be negative for the first time, and the machine is 'aware' that the system has been traversed completely. The control of the calculation is such that the machine continues to calculate and print any final results which may be required rather than to continue ray tracing.

Thus ray tracing is performed in a series of 'loops' of calculation, much of the programming technique consisting of arranging calculated data by shunting and other processes so that the nth cycle of any loop is operative on the nth cycle of numbers.

Several paraxial rays can be traced simultaneously in this manner, the time required for a single ray being about $\frac{1}{10}$ second per surface.

3.2. *First-Order Aberrations*

First-order aberrations are usually calculated either by tracing two paraxial rays, one from an axial object point, generally at infinity, the other through the diaphragm position, or by tracing the first ray and using the formulae given by Hopkins (1950). This latter method gives rise to the quantity $E=\bar{h}/hH$, where H is the Smith–Helmholtz invariant and $\bar{h}$ the semi-aperture of the diaphragm ray.

If A and B are the products of n and i for the axial and principal rays respectively, then the first-order aberrations are easily calculated in the form

$$S(1)=\sum A^2h\Delta\left(\frac{u}{n}\right),\quad S(2)=\sum ABh\Delta\left(\frac{u}{n}\right),\quad S(3)=\sum B^2h\Delta\left(\frac{u}{n}\right),$$

$$S(4)=\sum H^2P \text{ where } P=-c\Delta\left(\frac{1}{n}\right),\quad S(5)=\sum\frac{B}{A}\left[H^2P+B^2h\Delta\left(\frac{u}{n}\right)\right].$$

When A is zero $S(5)$ becomes infinite, and the form

$$S(5)_K=-H^2PBhE(2+AhE)-B^3h\Delta\left(\frac{1}{n^2}\right)$$

must be employed.

These formulae can be calculated with great rapidity on the machine, for having calculated $S(1)$, the values of $S(2)$ and $S(3)$ are in their respective storage locations awaiting output or summation within 7 msec. The complete first-order aberrations require 0·4 sec per surface. Intermediate values (i.e. separate surface sums or E values) can be either printed or punched out in decimal form or stored and summed during or at the end of the calculation.

3.3. *Finite Ray-Traces in the Meridian Section*

A ray-trace using the variables $\bar{v}$ (the length of the perpendicular from the pole of the surface to the ray) and $\sin U$, the angular inclination of the ray with the optical axis, has proved particularly easy to programme. This method was due originally to T. Smith. The refraction equations are

$$\bar{v}c=\sin U+\sin I,$$

$$\Delta(n\sin U)=c\bar{h}\Delta(n\cos I)=c\Delta(n\bar{v}),$$

where $\bar{h}$ is the height of incidence of the ray at the surface. Cosines are obtained from sines by a square root routine which is called in at each surface. The quantities $\bar{h}$ and $\bar{v}$ are connected by the relation

$$\bar{h}=\bar{v}\left[1-2\Big/\frac{(1+\cos U)(1+\cos I)}{\sin U\sin I}+1\right].$$

The transfer operation using these variables is very simple in that

$$(\bar{v})_2=(\bar{v}')_1-d\sin U_2$$

where d is the axial distance between the surfaces. This ray-trace is suitable for many reasons, one of them being that only two quantities are required as input data for each ray. Time can be saved by tracing several rays simultaneously; the reason for this is that every time a ray passes from one surface to another a magnetic instruction is required to bring the surface data concerned from the magnetic storage to the appropriate electronic store. This magnetic transfer requires 38 m sec and is comparable with the time required for some 17 multiplications or 30 additions or subtractions. A programme has been used on the Manchester machine by which three rays are traced at the same time, although

five can be traced using the magnetic storage purely for the retention of surface data. The tracing of, say, 25 rays at the same time would require magnetic instructions to retain calculated information during the tracing, so nothing is gained by tracing more than five rays simultaneously.

The time taken to trace three finite rays through one surface is enough to allow one to 'see' the rays in transit on the appropriate monitor tube, being about $1\frac{1}{2}$ seconds. As three rays are necessary for the calculation of axial spherical aberration at three apertures and for tangential coma at a particular field angle, the same three-ray routine can be used for both purposes. Separate smaller routines are necessary for the calculation and output of the values of these aberrations in a chosen focal plane. It is a fairly simple matter to modify the three finite meridional ray-traces to calculate the chromatic variation of spherical aberration; the refractive indices for the wavelengths chosen are supplied to the machine as extra surface data, and B lines are used to select the appropriate index from each set of surface data.

3.4. *Astigmatic Traces and Distortion*

Astigmatic ray-traces have been programmed as auxiliary routines to the meridional ray-traces already mentioned. When one surface has been traversed another routine is called in which calculates the astigmatism along the central (principal) ray of the three. Formulae have been described by Hopkins (1950) and Black (1953) using the paraxial angular variables h_S, u_S; h_T, u_T. These formulae are particularly well adapted to electronic calculation. Briefly they can be written

$$z\Delta(nu_S) = czh_S\Delta(n\cos I) = h_S\Delta(nu_T\cos I); \quad z = \frac{h_T}{\cos I} = \frac{h_T'}{\cos I'}$$

$$(h_S)_2 = (h_S)_1 - D(u_S)_2; \quad (h_T)_2 = (h_T')_1 - D(u_T)_2$$

where D is the distance along the ray between surfaces. The paraxial angular variables mentioned have been incorporated in new definitions of astigmatism, sagittal and tangential image curvature, which remain finite, determinate and meaningful in all cases. An account of the new definitions is given in the last-mentioned paper; they are: $\Delta(nuu_S\,\delta S)$ for sagittal image curvature, $\Delta(nuu_T\,\delta T)$ for tangential image curvature, $\Delta(nu_Su_T\,\delta A)$ for complete astigmatism, where δS and δT are the distances from the sagittal and tangential foci to the gaussian image plane. In the definition of astigmatism δA is the distance between the sagittal and tangential foci. Routines which trace three finite meridional rays and astigmatic traces along one of them require 2 seconds per surface.

The aberrations can be calculated in any form required, for instance as longitudinal intercepts or in the precise meaningful form

$$\sum_1^k \Delta(nuu_S\,\delta S); \quad \sum_1^k \Delta(nuu_T\,\delta T); \quad \sum_1^k \Delta(nu_Su_T\,\delta A),$$

where k is the number of surfaces. Results can be printed or punched out as before, either as surface contributions or as surface sums. We can define the precise distortion arising at a surface as $\Delta(nu\,\delta\bar{\eta})$ where $\delta\bar{\eta}$ is the difference in the heights in the paraxial image plane of a finite principal ray and the corresponding paraxial ray and we have $\Delta(nu\,\delta\bar{\eta}) = \Delta(nu\bar{\eta}) - \Delta(nu\eta)$. Now $\Delta(nu\eta) = 0$, and we have $\Delta(nu\,\delta\bar{\eta}) = \Delta(nu\bar{\eta})$. The calculation of the precise distortion contribution at each surface is a simple matter of a few instructions.

3.5. *Skew Rays*

Of the two best known methods of tracing skew rays, trigonometric or algebraic, the latter is best suited to electronic calculation. Any ray may be defined in terms of its direction cosines and the point of incidence, the general problem being that of determining the direction cosines of the refracted ray. If an incident ray has direction cosines L, M and N, the refracted ray has direction cosines L', M' and N', and the normal at the point of incidence has direction cosines A, B and C, then the coplanar refraction law can be expressed

$$\begin{vmatrix} L & M & N \\ L' & M' & N' \\ A & B & C \end{vmatrix} = 0.$$

The cosine of the angle of incidence is given by

$$\cos I = LA + MB + NC$$

and we have $\Delta(n \sin I) = 0$ and $L^2 + M^2 + N^2 = L'^2 + M'^2 + N'^2 = 1$. Thus the direction cosines of the refracted ray are determinable. The intersection point of the refracted ray with some further surface is determinable in general from the equation of the ray:

$$\frac{X-x}{L'} = \frac{Y-y}{M'} = \frac{Z-z}{N'} = D$$

and the position and form of the surface. Here D is the distance along the ray between surfaces and X, Y and Z; x, y and z are the coordinates of the known and unknown points respectively. For central quadrics of the general form

$$\alpha x^2 + \beta y^2 + \gamma z^2 = 1$$

the coordinates of the unknown intersection point are $(x, y, z) = X + L'D, Y + M'D, Z + N'D)$, and the parameter D is thus determined in general from the quadratic equation

$$(\alpha L'^2 + \beta M'^2 + \gamma N'^2)D^2 + 2(\alpha L'X + \beta M'Y + \gamma N'Z)D + \alpha X^2 + \beta Y^2 + \gamma Z^2 - 1 = 0.$$

The equation of the normal at the new point may be written

$$\frac{X-x}{\alpha x} = \frac{Y-y}{\beta y} = \frac{Z-z}{\gamma z}.$$

For paraboloids of the general form

$$\alpha x^2 + \beta y^2 + 2\gamma z = 0$$

the corresponding quadratic in D is

$$(\alpha L'^2 + \beta M'^2)D^2 + 2(\alpha L'X + \beta M'Y + \gamma N')D + (\alpha X^2 + \beta Y^2 + 2\gamma Z) = 0.$$

The equation of the normal at the new point is

$$\frac{X-x}{\alpha x} = \frac{Y-y}{\beta y} = \frac{Z-z}{\gamma}.$$

For cylindrical surfaces A or $B = 0$ when the plane of zero curvature lies in the sagittal or tangential section, and we have the two very useful relations

$$\Delta(n \cos I) = B\Delta(nM) + C\Delta(nN) \text{ for the sagittal section}$$

and

$$\Delta(n \cos I) = A\Delta(nL) + C\Delta(nN) \text{ for the tangential section.}$$

For spherical surfaces it is profitable to trace several skew rays at a time, and a series of routines have been used successfully which trace three skew rays simultaneously. The time required to trace three skew rays through one spherical surface is $1\frac{3}{4}$ seconds.

§4. Conclusion

Programmes and routines are now used on the Manchester electronic computing machine which trace all the rays normally used by the designer. These routines have been incorporated into programmes which calculate the first-order aberrations, the chromatic aberrations, spherical aberration at any number of apertures, tangential coma at any aperture or field angle, astigmatism at any number of field angles, oblique spherical aberration and distortion.

The determination of skew spherical aberration and skew coma presents no more difficulty than any of the aberrations in the meridian section, and as the time taken to calculate any set of aberrations is so small, much more extensive numerical investigations of lens system are possible. The magnetic storage of the machine which is the fundamental measure of capacity is so large as to accommodate most extensive amounts of information. Some of this space is being used in experiments in the automatic design of optical systems.

References

Black, G., 1953, *Thesis*, University of London.

Hopkins, H. H., 1950, *Wave Theory of Aberrations* (Oxford : Clarendon Press).

Kilburn, T., Toothill, G. C., Edwards, D. B. G., and Pollard, B. W., 1953, *Proc. Instn Elect. Engrs*, **100**, Pt. II, 487.

The Theory of Rectification and Injection at a Metal–Semiconductor Contact

By J. B. GUNN
Radar Research Establishment, St. Andrew's Road, Gt. Malvern, Worcs.

Communicated by R. A. Smith; MS. received 10th March 1954

Abstract. The diffusion equations for holes and electrons are solved for the region of a semiconductor near a planar contact with a metal, using assumptions appropriate to transistor-quality silicon and germanium. The hole and electron currents are calculated separately, and from them are obtained the rectification equation and injection ratio of the contact.

§ 1. Introduction

Numerous quantitative theories have been given for the relation between current density and applied voltage at a metal–semiconductor contact (e.g. Schottky 1939, Bethe 1942) but these have all adopted what may be called the *pre-transistor* approach, in which the (usually implicit) assumption is made that the minority carrier density does not depart sensibly from its equilibrium value. This is equivalent to the assumption that the minority carrier lifetime is very short, and is plainly unjustified in the case of transistor materials. In the *post-transistor* approach, this restriction is not made, and entirely different results may be expected. The appropriate methods have been demonstrated by Shockley (1949) in his analysis of p–n junctions, but do not yet appear to have been applied to the practically important case of the contact between a metal and a transistor semiconductor.

In this paper an attempt is made to treat the metal–semiconductor contact by these methods, but it should be stressed at the outset that the conditions at a point contact are so different from those assumed here that the theory may be hardly, if at all, applicable to this case.

§ 2. Formulation and Assumptions

In formulating the problem, we shall assume that the potential distribution near the metal–semiconductor contact has the classical form shown in figure 1.

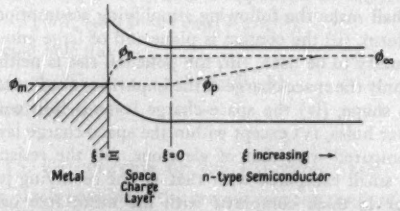


Figure 1. Potential distribution and quasi-Fermi levels near the contact.

(Henceforth only an n-type semiconductor is considered, for the sake of definiteness ; the analysis applies, with obvious changes, equally to p-type.) A region of positive space charge near the surface produces a potential rise (the term 'barrier' is intentionally avoided) whose magnitude is determined by the surface acceptor levels, which are assumed to define the position of the Fermi level at the surface. At equilibrium, the Fermi levels in the metal, at the surface, and deep in the semiconductor are, of course, equal; when a voltage is applied, the level in the semiconductor is displaced with respect to the other two. The resulting changes in carrier distribution give rise to currents of both holes and electrons, which may, under certain simplifying assumptions, be calculated separately. The rectification characteristic is then obtained by adding the two currents, while the injection efficiency is obtained from their ratio.

In order to simplify the analysis, it will be convenient to express the problem in terms of the following dimensionless quantities:

$\nu = n/n_i$, $\rho = p/n_i$, $\Omega = N_0/n_i$, $\psi = qP/kT$, ξ, η, $\zeta = x/L_i$, y/L_i, z/L_i, $\tau = t/t_i$, $\Pi = t_i R/n_i$, $\boldsymbol{\nu} = \mathbf{n} \cdot t_i/n_i L_i$, $\boldsymbol{\rho} = \mathbf{p} \cdot t_i/n_i L_i$.

Here n_i = electron density in intrinsic material, $q = |$electronic charge$|$, k = Boltzmann's constant, T = absolute temperature, n = electron density, p = hole density, N_0 = net impurity density (taken negative in n-type material), x, y, z are Cartesian coordinates, t = time, R = net rate of generation of hole–electron pairs, P = electrostatic potential, $\mathbf{n}$ = electron flux = $-$(electron current)/q, $\mathbf{p}$ = hole flux = hole current/q. The characteristic length L_i is given by

$$L_i = (kTK/4\pi q^2 n_i)^{1/2} \quad (K = \text{dielectric constant})$$

and the characteristic time by $t_i = L_i^2 q/(kT\mu)$ (μ = hole mobility). These are related to the Debye space–charge length and the intrinsic dielectric relaxation time respectively.

In discussing non-equilibrium situations we shall, following Shockley, introduce the quasi–Fermi levels ϕ_n and ϕ_p, defined by $\nu = \exp(\psi - \phi_n)$, $\rho = \exp(\phi_p - \psi)$.

The standard equations governing the flow of holes and electrons can now be expressed as follows:

$$\boldsymbol{\nu} = b\nu\nabla\psi - b\nabla\nu \quad \ldots\ldots(1) \qquad \boldsymbol{\rho} = -\rho\nabla\psi - \nabla\rho \quad \ldots\ldots(2)$$

$$\nabla \cdot \boldsymbol{\nu} = -\partial\nu/\partial\tau + \Pi \quad \ldots\ldots(3) \qquad \nabla \cdot \boldsymbol{\rho} = -\partial\rho/\partial\tau + \Pi \quad \ldots\ldots(4)$$

$$-\nabla^2\psi = \rho - \nu - \Omega \quad \ldots\ldots(5)$$

or in the alternative forms

$$\boldsymbol{\nu} = b\nu\nabla\phi_n, \quad \boldsymbol{\rho} = \rho\nabla\phi_p \quad \ldots\ldots(1a), (2a)$$

(b = electron mobility/hole mobility).

Finally, we shall make the following simplifying assumptions: (i) the semiconductor is uniform, (ii) the contact is plane and of large enough area for one-dimensional geometry to be used, (iii) the potential rise is neither very high nor very low, so that only the space charge of the impurities need be taken into account in calculating its shape, (iv) the space-charge layer is thin compared with the diffusion length for holes, (v) except within the space-charge layer the density of holes is small compared with that of electrons, (vi) the resistance of the bulk semiconductor is small compared with that of the rectifying junction, (vii) the space-charge layer is thick compared with the mean free path of electrons, (viii) the field in the space-charge layer is not high enough to cause a change in

mobility (Shockley 1951), (ix) only the steady state is considered, (x) the origin of ξ is taken at the inner side of the space-charge layer, (xi) the surface states are very dense compared with the number of ions in the space-charge layer.

§3. Boundary Conditions for ϕ_n and ϕ_p

Since it is a property of a metal that it permits holes and electrons to recombine instantaneously, within it both ϕ_p and ϕ_n must be equal to ϕ_m, the Fermi energy. At a sufficiently great distance into the semiconductor, the hole density will have decayed to ρ_∞, its equilibrium value, the potential will be ψ_∞, and ϕ_p and ϕ_n will again be equal, their value being denoted by ϕ_∞. If Φ is the voltage applied across the rectifier, then

$$\Phi = \phi_m - \phi_\infty . \qquad \ldots\ldots(6)$$

The arguments given by Shockley can be applied also to this case to show that, provided assumptions (iv), (v) and (vi) are justified, ϕ_n and ψ must be constant for ξ greater than 0, and ϕ_p constant for ξ less than 0.

Hence ϕ_{n0}, ϕ_{p0}, and ψ_0 the values at $\xi = 0$, must be given by

$$\phi_{n0} = \phi_\infty , \quad \phi_{p0} = \phi_m , \quad \psi_0 = \psi_\infty . \qquad \ldots\ldots(7),(8),(9)$$

§4. Calculation of the Hole Flux

In a rigorous calculation, both ν and ρ would be obtained by treating equations (1)–(5) as a system of simultaneous differential equations. It is a consequence of assumption (v), however, that, for ξ greater than 0, ρ can be calculated to a good approximation from equations (1) and (2) alone, the electrons being supposed only to compensate the space-charge of the holes and to be otherwise without effect. Assumption (vi) permits ψ to be taken as constant for ξ greater than 0, so that the equation for ρ reads, for the steady state :

$$-\nabla^2\rho = -\frac{d^2\rho}{d\xi^2} = \Pi . \qquad \ldots\ldots(10)$$

Again, subject to assumption (v), Π may be expressed in terms of the hole lifetime τ_p (in reduced units)

$$\Pi = (\rho - \rho_\infty)/\tau_p . \qquad \ldots\ldots(11)$$

The solution of (10) and (11) which satisfies the boundary conditions

$$\rho - \rho_\infty \to 0 \qquad \text{as} \quad \xi \to \infty \qquad \ldots\ldots(12)$$

and

$$\rho = \exp(\phi_{p0} - \psi_0) \quad \text{at} \quad \xi = 0 \qquad \ldots\ldots(13)$$

with the substitutions of (6) and (8), is given by

$$\rho - \rho_\infty = \rho_\infty(\exp\Phi - 1)\exp(-\xi/\tau_p^{1/2}). \qquad \ldots\ldots(14)$$

Hence ρ_0, the injected hole flux, is

$$\rho_0 = (-d\rho/d\xi)_{\xi=0} = \rho_\infty \tau_p^{-1/2}(\exp\Phi - 1) \qquad \ldots\ldots(15)$$

which is the required equation.

§5. Calculation of the Electron Flux

The calculation of ν is possible in the present case as a result of assumption (iii), which permits the calculation of $\psi(\xi)$ in the space-charge layer independently of both ν and ρ. Since both the latter are neglected, we have $d^2\psi/d\xi^2 = \Omega$.

A solution satisfying the boundary conditions $d\psi/d\xi = 0$, $\psi = \psi_0$ at $\xi = 0$ is given by

$$\psi = \psi_\infty + \tfrac{1}{2}\Omega\xi^2 \qquad \ldots\ldots(16)$$

on substituting from equation (9).

Since recombination within the space-charge layer is to be neglected, equation (3) becomes

$$\nabla \cdot \mathbf{v} = 0 \qquad \ldots\ldots(3\,a)$$

and $\mathbf{v} = \mathbf{v}_0 =$ constant in the space-charge layer. Equation (1 *a*) can be rewritten as

$$b \exp(\psi - \phi_n) \frac{\delta\phi_n}{\delta\xi} - \mathbf{v}_0 = 0$$

which is a differential equation for ϕ_n. On substituting equation (16) and separating variables, we obtain

$$\exp(-\phi_n)\, d\phi_n - (\mathbf{v}_0/b) \exp(-\psi_\infty) \exp(-\Omega\xi^2/2)\, d\xi = 0. \qquad \ldots\ldots(17)$$

This can be integrated in terms of the function $E(u)$, defined by

$$E(u) = \int_0^u \exp(w^2)\, dw. \qquad \ldots\ldots(18)$$

The solution of (17) is then

$$-\exp(-\phi_n) - (-2/\Omega)^{1/2} (\mathbf{v}_0/b) \exp(-\psi_\infty) \cdot E\{(-\Omega/2)^{1/2}\xi\} + c = 0. \qquad \ldots\ldots(19)$$

The values of c and $\mathbf{v}_0$ are determined by the boundary conditions

$\phi_n = \phi_\infty$ at $\xi = 0$

$\phi_n = \phi_m$ at $\xi = \Xi$, the value of ξ at the surface.

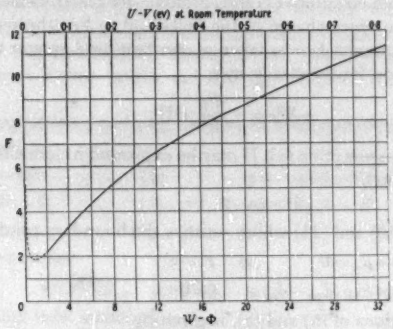


Figure 2. The function $F(\Psi - \Phi)$.

Thus, on making use of (6), and putting $\nu_\infty = \exp(\psi_\infty - \phi_\infty)$, equation (17) leads to

$$\mathbf{v}_0 = \frac{b\,\nu_\infty(-\Omega/2)^{1/2}\{1 - \exp(-\Phi)\}}{E\{(-\Omega/2)^{1/2}\Xi\}}. \qquad \ldots\ldots(20)$$

Assumption (xii), that the surface states are dense, implies that $\phi_m - \psi_\Xi$ ($\psi_\Xi = \psi$ at $\xi = \Xi$) is constant independent of V. Hence, if Ψ is the equilibrium value of $\psi_\infty - \psi_\Xi$, the relation $\psi_\infty - \psi_\Xi = \Psi - \Phi$ permits Ξ to be expressed by

$$(-\Omega/2)^{1/2}\Xi = (\Psi - \Phi)^{1/2}.$$

Now the function $F(\Psi-\Phi)\equiv\exp(\Psi-\Phi)/E\{(\Psi-\Phi)^{1/2}\}$ varies only slowly with $\Psi-\Phi$ (see figure 2) so that equation (20) may be more conveniently expressed as

$$\nu_0=b\nu_\infty(-\Omega/2)^{1/2}\exp(-\Psi)\ F(\Psi-\Phi)(\exp\Phi-1) \qquad \ldots\ldots(21)$$

which is the required expression for ν_0. It should be noted that the negative sign is to be taken for $(-\Omega/2)^{1/2}$.

§ 6. Discussion

Since the total current density i through the contact is given by

$$i/q=\rho_0-\nu_0=(\rho_s-\nu_s)(\exp\Phi-1),$$

the contact may be regarded as made up of two rectifiers in parallel, one conducting only by holes and the other only by electrons. In extrinsic material, ρ_∞ and ν_∞ may be replaced by $-\Omega^{-1}$ and $-\Omega$ respectively. Hence the saturation current ρ_s of the hole rectifier is $-1/\Omega\tau_p^{1/2}$. The corresponding term ν_s for the electron rectifier is not properly a saturation current, since it depends on Φ, but it varies only slowly and may be termed the 'quasi-saturation current', given by

$$\nu_s=(b/\sqrt{2})(-\Omega)^{3/2}\exp(-\Psi)\ F(\Psi-\Phi). \qquad \ldots\ldots(22)$$

If ρ_s is at least comparable with ν_s, the contact will act as a hole emitter, the injection ratio γ being given by $\gamma=\theta/(\theta+1)$ where

$$-\theta\equiv\rho_s/\nu_s=(\sqrt{2}/b)(-\Omega)^{-5/2}\tau_p^{-1/2}\exp\Psi\ F(\Psi-\Phi). \qquad \ldots\ldots(23)$$

A knowledge of the relation between γ and Ψ is of importance since, as has been pointed out (Hogarth 1953), values of Ψ calculated from measurements of γ may be more reliable than those obtained in other ways. In calculating this relation, Hogarth obtained an equation corresponding to (23) which, with the present notation and assumptions, may be written as

$$\theta'=(-\Omega)^{-2}\exp\Psi. \qquad \ldots\ldots(24)$$

The assumptions used in the derivation of this equation are effectively those of the pre-transistor approach, however, since the hole current is calculated from the rate at which holes can cross the 'barrier', and not from the rate at which they can diffuse away from it. Equations (23) and (24) differ by a factor of $\theta'/\theta=(b/\sqrt{2})(-\Omega\tau_p)^{1/2}/F$; for specimens typical of those used by Hogarth, $\Omega\sim-10$, $\tau_p\sim10^5$, $F\sim10$, so that $\theta'/\theta\sim10^2$. Values of Ψ calculated from (24) will be lower than those from (23) by an amount $\Delta\Psi=\ln\{(b/\sqrt{2})(-\Omega\tau_p)^{1/2}/F\}\sim4{\cdot}6$ $=0{\cdot}12$ ev at room temperature. However, it seems doubtful whether more than qualitative conclusions can be drawn from experiments with point contacts.

The value of γ at a metal–semiconductor contact has also been discussed by Schwartz and Walsh (1953), who obtain an expression equivalent to (15) for the hole current. For the electron current, however, they take the classical diode theory formula, which may be written $\nu_s=-C\Omega\exp(-\Psi)$, where $C=(t_i/L_i)(kT/2m\pi)^{1/2}=5{\cdot}2$ at room temperature. Thus the current ratio is

$$\theta''=C^{-1}(-\Omega)^{-2}\tau_p^{-1/2}\exp\Psi. \qquad \ldots\ldots(25)$$

For the same parameters as before, the factor $\theta''/\theta=b(-\Omega)^{1/2}/(\sqrt{2}\,CF)$ is about 10^{-1} and the calculated value of Ψ higher by about $2{\cdot}3=0{\cdot}06$ ev than that obtained from (23). These authors' experimental conditions would seem to fulfil the requirements of the present theory rather better than those of the diode theory, since the thickness of the space-charge layer must be several mean free paths (see below). Equations (23) is consequently to be preferred to equation (25).

In this connection it is interesting to observe that, no matter what model is taken for the rectification mechanism, θ is unlikely to vary rapidly with V. Hence, if measurements of γ have shown that, as for example in germanium diodes, θ is large for forward voltages, it would be expected to be large also for small inverse voltages. The temperature dependence of the inverse current should thus be that of ρ_s, which is independent of Υ. Even under the very different conditions obtaining at a point contact, we may expect that, measured in conventional units, ρ_s will be approximately proportional to n_i^2, and so should increase as $\exp(-E_g/kT)$ with the energy gap E_g equal to 0·72 ev in germanium. It is consequently hardly surprising that the deduction of the 'barrier height' in germanium diodes from the temperature dependence of the inverse current should give results near this value. The differences between E_g and the observed values for the activation energy (0·6–0·7 ev) may be perhaps ascribed to changes in τ_p with temperature, and to the contact geometry.

In conclusion, some of the assumptions of § 2 may be compared with practical conditions in silicon and germanium. The requirements of (i) are easily met, only formed contacts being excluded, but those of (ii) and (vi) are probably stringent enough to exclude point contact rectifiers from comparison with the theory. Simple calculation shows that in these the radius of the space-charge layer may be many times that of the contact area, so that the geometry is not even approximately one-dimensional. Again, where $\nabla\psi$ is not neglected, the nature of the carrier flow is altered, and quite different results may be expected. Assumption (iv), on the other hand, is an excellent approximation for material of the commonly used qualities, as is assumption (v) when Φ is not much greater than 1. Assumptions (vii) and (viii) are probably of marginal validity, since for germanium with $\Omega = -10$, $\Upsilon = 20$, we have $\Xi = -2$, and $(\delta\psi/\delta\xi)_\Xi = 20$, compared respectively with values of approximately 0·1 for the mean free path and 40 for the field at which the electron drift velocity becomes constant (Ryder 1953). No quantitative discussion of the surface properties of semiconductors is possible without assumption (xi), or some similar assumption relating $\phi_m - \psi_\Xi$ to some other variable. It must be admitted, however, that there is little direct evidence to support it.

§ 7. Conclusion

The theory presented, which is the logical extension of the method of quasi-Fermi levels to the case of a metal–semiconductor contact, permits the calculation of the current–voltage characteristic and the injection ratio in terms of the impurity content and hole lifetime of the bulk material and the height of the surface potential rise. The analysis is limited in its application by the assumptions made, and applies only qualitatively to point contacts. It should, however, apply to area contacts of the types described by Granville and Henisch (1952), Schwartz and Walsh (1953) and Gunn (1954).

Acknowledgments

This paper is published by permission of the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office. I am indebted to Mrs. A. M. Woodward for the calculation of the function F.

References

Bethe, H. A., 1942, *Massachusetts Institute of Technology Radiation Laboratory Report*, **43**, 12.

Granville, J. W., and Henisch, H. K., 1952, *Proc. Phys. Soc.* B, **65**, 650.

Gunn, J. B., 1954, *Proc. Phys. Soc.* B, **67**, 409.

Hogarth, C. A., 1953, *Proc. Phys. Soc.* B, **66**, 845.

Ryder, E. J., 1953, *Phys. Rev.*, **90**, 766.

Schottky, 1939, *Z. Phys.*, **113**, 367.

Shockley, W., 1949, *Bell. Syst. Tech. J.*, **28**, 435; 1951, *Ibid.*, **30**, 990.

Schwartz, R. F., and Walsh, J. F., 1953, *Proc. Inst. Radio Engrs, N.Y.*, **41**, 1715.

APPENDIX I

Values of the Units used

In germanium and silicon, the reduced units used in the analysis have the following values at room temperature:

Symbol	Unit of:	Value in Ge	Value in Si
n_i	Carrier density (cm^{-3})	$2{\cdot}4\times10^{13}$	$6{\cdot}8\times10^{10}$
kT/q	Potential (v)	$2{\cdot}5\times10^{-2}$	$2{\cdot}5\times10^{-2}$
L_i	Length (cm)	$9{\cdot}6\times10^{-5}$	$1{\cdot}58\times10^{-3}$
t_i	Time (sec)	$1{\cdot}84\times10^{-10}$	4×10^{-7}
kT/qL_i	Electric field (v cm^{-1})	260	15·8
n_iL_i/t_i	Carrier flux ($cm^{-2}sec^{-1}$)	$1{\cdot}25\times10^{19}$	$2{\cdot}66\times10^{14}$
	Corresp. current density (A cm^{-2})	2·0	4·25
b	Mobility ratio	2·1	5

APPENDIX II

Results Expressed in Conventional Units

In the following equations, $\mathbf{p}_0$ corresponds to ρ_0, p_∞ to ρ_∞, t_p to τ_p, V to Φ, $\mathbf{n}_0$ to ν_0, n_∞ to ν_∞, and U to Ψ.

Equations (15), (21) and (23), for the hole and electron currents and the current ratio θ, may be transformed into the following, where the quantities are expressed in conventional units:

$$\mathbf{p}_0 = p_\infty \left(\frac{kT\mu}{qt_p}\right)^{1/2} \{\exp(qV/kT)-1\} \qquad \ldots\ldots(15\,A)$$

$$\mathbf{n}_0 = b\mu n_\infty \left(\frac{-2\pi kTN_0}{K}\right)^{1/2} \exp\left(-\frac{qU}{kT}\right) F\left\{\frac{q}{kT}(U-V)\right\}\left\{\exp\left(\frac{qV}{kT}\right)-1\right\} \qquad \ldots\ldots(30\,A)$$

$$\theta = -\left(\frac{K}{2\pi\mu q b^2}\right)^{1/2} n_i^2(-N_0)^{-5/2} t_p^{-1/2} \exp\left(\frac{qU}{kT}\right) F\left\{\frac{q}{kT}(U-V)\right\} \qquad \ldots\ldots(35\,A)$$

RESEARCH NOTES

A Comparison of the Rates of Change of Current in the Step and Return Processes of Lightning Flashes

By D. B. HODGES
The University of Natal, Durban, South Africa

Communicated by B. F. J. Schonland; *MS. received* 22*nd March* 1954

1. THE electrical field-change due to a lightning discharge at a distance exceeding 50 km is almost entirely due to the radiation component of the field and consists of rapid pulsations. The relative amplitudes of these pulsations depend upon the rates of change of current in the channel producing them and will, under certain conditions, be proportional to these rates of change. These conditions are: (*a*) the channel elements to be compared must be parallel, (*b*) the field-change recording system must have the same amplification for the pulsations to be compared, which may differ in wave-form, and (*c*) the distance must be sufficiently great for conditions of propagation to be the same and yet not so great that ionospheric reflections can confuse the measurement of amplitudes.

2. Such measurements, as Schonland (1953) has shown, can be used to estimate the ratio of the maximum rate of change of current (dI/dt) in the first return streamer to the average value of the rate of change of current in the sudden step portions of the stepped leader preceding it (di/dt). They have been used by him to derive the magnitude of the current changes in the step process, since the average value of dI/dt is known from direct measurements on lightning flashes to tall structures, and the duration of the sudden step is approximately known.

The present note is written to make available the data used in this discussion.

3. The ratios of the average amplitudes $\bar{p}$ of a stepped leader ripple to the amplitudes P of the succeeding return streamer radiation fields were obtained from a number of oscillograph records of radiation wave forms made in past years at the Bernard Price Institute, Johannesburg, by various workers.

The records were separated for analysis into two classes, corresponding to type α and type β leaders. The distinguishing characteristic of the field change from a type α leader is that the pulsations persist with a fairly uniform though relatively small amplitude over the whole duration of this leader. The radiation from a type β leader, on the other hand, shows a first stage of short duration ($\sim$1·3 msec) during which the amplitude of the ripple is relatively large, followed by a second and longer stage of α type (duration $\sim$9 msec) in which the ripple is barely observable.

Figures 1 and 2 are histograms of $P/\bar{p}$ for type α and type β leaders respectively.

Since an approximate upper limit for $\bar{p}$ was all that was required, records which gave very small type α leader ripple, readable with difficulty, were often

rejected. For the same reason no particular attention was paid to occasional exceptionally small amplitudes in the leader ripple. This is equivalent to ignoring those steps which have either a large horizontal component in direction or involve unusually small values of di/dt. The first of these two possibilities means that condition (*a*) of § 1 is approximately satisfied, since the selected steps will be approximately vertical, as is the bottom of the return streamer channel which is responsible for the peak value of dI/dt. Condition (*b*) was satisfied since the amplifying system was well able to deal faithfully with the pulsations of leader and return-streamer, which both have half-periods of the order of 20 μsec. There was no difficulty in selecting records which satisfied condition (*c*).

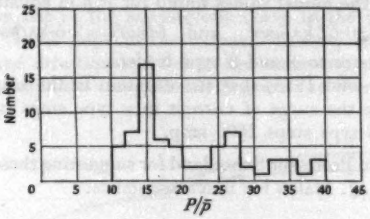


Figure 1. Distribution of values of $P/\bar{p}$ for α type leaders.

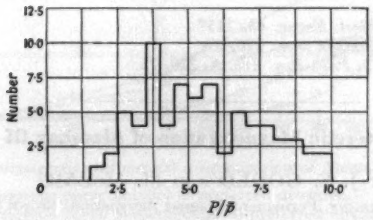


Figure 2. Distribution of values of $P/\bar{p}$ for β type leaders.

4. From figure 1 it is seen that for type α leaders $P/\bar{p}$ ranges from 10 to 40 with a modal value of 15. These leaders were of most interest in Schonland's discussion of the stepped leader process, since he was concerned with steps of moderate length (~20 m).

Type β leaders, whose step lengths may be as long as 200 m and which show up very brightly on camera records, give values of $P/\bar{p}$ ranging from 2 to 10 with a modal value of 3·5.

5. As has been pointed out, we may write

$$P/\bar{p} = (dI/dt) \Big/ di/dt$$

or

$$di/dt = \frac{\bar{p}\, dI/dt}{P}. \qquad \ldots\ldots (1)$$

Observed values of dI/dt are shown in the following table.

Rate of Rise of Current Peaks in the Return Stroke

Object struck	Reported by	Range (kA/μsec)	Model range (kA/μsec)
Empire State Building	Hagenguth and Anderson (1952)	0–50	12·5*
Tall structures	McCann (1944)	0–45	9·0*
Transmission line towers	Berger (1936)	0–31	11·0†
Ground (by frame) aerial induced e.m.f.	Norinder (1937)	0–30	4·5†

* Smoothed or effective dI/dt. † max dI/dt

The mean of the modal values of dI/dt in the last column is 9·25 kA/μ sec. Substituting this, and the modal values found for $P/\bar{p}$ in equation (1),

$$(di/dt)_\alpha = 0{\cdot}62 \text{ kA}/\mu\text{sec}, \quad \text{and} \quad (di/dt)_\beta = 2{\cdot}6 \text{ kA}/\mu\text{sec}$$

where the subscripts refer to α and β type leaders.

It is known (Schonland 1953) that the duration of the bright step process is 1 μsec or less, hence the surge of current in α type steps is of the order of 620 amp and that in β type steps 2600 amp.

My thanks are due to Professor Schonland for suggesting these measurements, and to him and Dr. D. J. Malan for much assistance.

References

Berger, K., 1936, *Ass. Suisse des Elect., Bull.*, **27**, No. 6, 145.
Hagenguth, J. H., and Anderson, J. G., 1952, *Amer. Instn Elect. Engrs*, Tech. Paper, 52–175.
McCann, G. D., 1944, *Elect. Engng*, **63**, 1157.
Norinder, H., 1935, *J. Franklin Inst.*, **220**, 69.
Schonland, B. F. J., 1953, *Proc. Roy. Soc.* A, **220**, 25.

Anhysteretic Magnetization of Alcomax III

By J. E. Gould and M. McCaig
Central Research Laboratory, Permanent Magnet Association, Brown Street, Sheffield

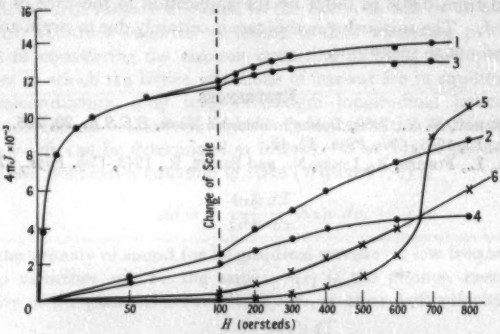
MS. received 26th March 1954

Anhysteretic magnetization of a ferromagnetic material is the magnetization J resulting from the application of a steady field H and a superposed large alternating field which is reduced gradually to zero. A plot of magnetization against steady field is known as the anhysteretic or 'ideal' magnetization curve and rises much more steeply than the normal magnetization curve. If the specimen has no external demagnetizing factor then the initial slope J/H of the anhysteretic curve is the anhysteretic susceptibility k_0. The reciprocal $1/k_0$ is sometimes known as the internal demagnetizing factor N_i. Néel (1943) has given a simple explanation of the anhysteretic phenomena, which are also discussed by Néel, Forrer, Janet and Baffie (1943).

Bulgakov (1950) has published measurements of N_i parallel and perpendicular to the preferred axis for Magnico (Russian Alcomax) and finds N_i about 20 times less in the preferred direction than in the perpendicular direction. Measurements have now been made on Alcomax III.

The tests were carried out on a bar 3·6 cm × 1·2 cm × 3·9 cm on the B.S.I. magnet tester (B.S. 406, 1931, type B). The alternating field was produced by a coil of 70–100 turns wound directly on the specimen and fitting inside the steady field coil. A 50 c/s current of up to about 50 amp r.m.s., obtained from a continuously variable transformer, could be passed through the inner coil. The maximum alternating current was limited by the inductance of the winding, which varied considerably according to the mean magnetization of the sample, but the peak value of the alternating field was always 1500 Oe or more.

The figure shows the anhysteretic magnetization curves in the preferred and transverse directions, the remanent magnetization on reducing the steady field to zero for both directions, and also the normal magnetization curves for both directions. Note that in this figure the H scale changes at 100 Oe, so that the very steep rise of the anhysteretic curve in the preferred direction is clear. From the initial slope of the curves, the anhysteretic susceptibility is 103 in the preferred direction and 1·9 in the transverse direction, whereas for the normal curves the initial susceptibility is greater in the transverse direction.



Anhysteretic and normal magnetization of Alcomax III.

1. Anhysteretic curve parallel to preferred direction.
2. Anhysteretic curve perpendicular to preferred direction.
3. Anhysteretic remanence curve parallel to preferred direction.
4. Anhysteretic remanence curve perpendicular to preferred direction.
5. Magnetization curve parallel to preferred direction.
6. Magnetization curve perpendicular to preferred direction.

Tests on the effect of varying values of the peak alternating field show that, for a fixed value of steady field, there is very little change in magnetization between 1000 and 1500 Oe peak field. At some lower value of peak field there is a very steep drop in magnetization, but the knee of the curve always occurs at approximately the same value of total field, i.e. steady field plus peak alternating field.

There are several practical consequences of these results.

The remanence curve demonstrates the possibility of magnetizing Alcomax by a combination of a direct field of 400–500 Oe in conjunction with an alternating field of 1500 Oe peak. For loop type magnets, often magnetized by one conductor, the cost of equipment to give high alternating currents is much less than for equally high direct currents.

The frequent difficulty in demagnetizing Alcomax may be partly due to anhysteretic effects caused by small stray fields. The earth's vertical field would be capable of leaving a remanent magnetization of 650 gauss in the preferred direction. A large magnet which is not wholly within the strong alternating field would be difficult to demagnetize because of its own stray field. Such effects are very much less if both alternating field and any residual stray field are in the transverse direction than if both fields are in the preferred direction.

The internal demagnetizing factors are 0·01 and 0·5 in the preferred and transverse directions against values found by Bulgakov (1950) of 0·02 and 0·4. The term 'internal demagnetizing factor' implies the effect of internal air gaps due to non-magnetic inclusions. The difference in the values found in the two directions suggests a connection between this difference and the evidence which has been found for a definite orientation of submicroscopic precipitated plates in field-cooled Alcomax. Although such a connection may exist it cannot be said that these measurements confirm the presence of oriented inclusions. It has been found that for a soft nickel wire the internal demagnetization factor is increased from 0·008 to 0·024 by the application to the wire of a tension of 5·2 kg mm^{-2}. The anisotropy in this case is certainly due to strain and magnetostriction and not to inclusions.

References

Bulgakov, N. V., 1950, *Doklady Akademii Nauk, S.S.S.R.*, **70**, 205.
Néel, L., 1943, *Cah. Phys.*, **17**, 47.
Néel, L., Forrer, R., Janet, N., and Baffie, R., 1943, *Cah. Phys.*, **17**, 51.

LETTERS TO THE EDITOR

The Effects of Heat Flow on Thermoelectric Power in Semiconductors

The effects on thermoelectric power of the conduction of heat by lattice vibrations have been discussed by Gurevich (1945) and Frederikse (1953). Both of these writers treat the matter somewhat formally, and the present author would like to present a method based on a more physical approach.

The conventional theory (Seitz 1940) of electron scattering by lattice vibrations shows that it is isotropic. This is because the phonons are assumed to be in thermal equilibrium, which will not be the case in experiments on thermoelectric power. They will be drifting down the temperature gradient and because of this the electron scattering will show this preferred direction. The additional drift velocity imparted to the electrons will be an average of the phonon drift velocities weighted according to their scattering power. This is equivalent to considering the electron scattering as being isotropic in a frame of reference in which the lattice vibrations of interest are in equilibrium.

In semiconductors, only long-wavelength longitudinal lattice vibrations need to be considered. The drift velocity of phonons of this kind in the frequency range ν to $\nu+d\nu$ can be determined as follows. The flux of phonons obtained by solving the Boltzmann equation of state (Wilson 1953) is

$$dJ = -\frac{4\pi \boldsymbol{k}}{3s^3\boldsymbol{h}}\frac{dT}{dz}\lambda(\nu)\nu\, d\nu.$$

Here s is the velocity of sound for longitudinal waves : at low frequencies, phase and group velocities will be the same; $\lambda(\nu)$ is the phonon mean free path. The density of the phonons is $(4\pi \boldsymbol{k}T/s^3\boldsymbol{h})\nu\, d\nu$, so their drift velocity must be

$$v(\nu) = -\frac{\lambda(\nu)s}{3T}\frac{dT}{dz}.$$

We then define an average drift velocity V_D by postulating an average mean free path λ^* for the phonons of importance.

If only lattice scattering needs to be considered, then the next step is to add a term $-n\boldsymbol{e}V_D$ to the expression for the electron current, where n is the density of electrons. Where impurity scattering is not negligible, this term has to be multiplied by l/l_T, where l is the total electron mean free path and l_T that due to phonon scattering only. The additional term is

$$\Delta I = \frac{n\boldsymbol{e}\lambda^* s}{3T}\frac{l}{l_T}\frac{dT}{dz}$$

and so the extra thermoelectric power from this cause is

$$\Delta Q = -\frac{(2\pi \boldsymbol{m}^*\boldsymbol{k})^{1/2}}{4\boldsymbol{e}}\frac{\lambda^* s}{l_T T^{1/2}}.$$

Due to the approximations in this treatment the result is a third of that derived by Frederikse.

The lattice thermal conductivity is $\kappa = \frac{1}{3}C_v s\bar{\lambda}$. Here, $\bar{\lambda}$ is the average mean free path of the phonons weighted according to their energies, so that it refers

mainly to phonons of the highest frequencies. Hence λ^* and $\bar{\lambda}$ will not be the same since mean free path decreases with increasing frequency.

Using Frederikse's value of κ for his specimen B, Estermann and Weertman's (1952) results for specific heat, and assuming (Shockley 1950) $s = 5 \times 10^5$ cm sec^{-1} at all temperatures, it was possible to calculate $\bar{\lambda}$. The electron mean free path l_T was calculated from Prince's (1953) data, and thence $\bar{\lambda}/l_T$ and ΔQ were obtained (see table). It will be seen that these results are about 100 times too small. This can be explained on the following considerations. The electrons will be scattered mainly by lattice vibrations whose wavelength is about half of the electron wavelength. Hence, at 30°K, the frequency of the lattice vibrations which interact with the electrons must be about 2×10^{11} sec^{-1} compared with the maximum frequency of about 8×10^{12} sec^{-1} calculated from the Debye temperature θ for germanium (370°K), using $k\theta = h\nu_{max}$. Measurement of ultrasonic absorption at about 3×10^7 sec^{-1} by Truell and Bronzo (1953) indicates that the mean free path is approximately inversely proportional to frequency. Assuming a similar relation at higher frequencies, this would suggest that λ^* would be approximately 40 times $\bar{\lambda}$, which removes most of the discrepancy. However, there are a number of complicating factors, e.g. $\bar{\lambda}$ refers to all three modes of vibration, whilst λ^* refers to the longitudinal mode only, which will obscure this apparently quite good agreement.

T (°K)	20	50	60	70	80
κ (W cm^{-1} deg^{-1})	11·5	7·5	5·1	4·2	3·5
C_v (joule cm^{-3} deg^{-1})	0·064	0·43	0·55	0·66	0·77
$\bar{\lambda}$ (10^{-5} cm)	108	10·5	5·6	3·8	2·7
l_T (10^{-5} cm)	20·6	8·24	6·87	5·89	5·16
$\bar{\lambda}/l_T$	5·24	1·27	0·816	0·645	0·523
ΔQ_{theor} (10^{-6} v deg^{-1})	−88	−13·5	−7·9	−5·8	−4·4
ΔQ_{exp} (10^{-3} v deg^{-1})	−8·0	−2·5	−2·0	−1·6	−1·3

I should like to thank Dr. T. E. Allibone for permission to publish this letter, and Dr. E. Billig and my colleagues for their helpful comments.

Research Laboratory, J. E. Parrott.
Associated Electrical Industries, Ltd.,
Aldermaston Court,
Aldermaston, Berks.
19th January 1954.

Estermann, I., and Weertman, J. R., 1952, *J. Chem. Phys.*, **20**, 972.
Frederikse, H. P. R., 1953, *Phys. Rev.*, **92**, 248.
Gurevich, L., 1945, *J. Phys., Moscow*, **9**, 477.
Prince, M. B., 1953, *Bull. Amer. Phys. Soc.*, **28**/2, 10.
Seitz, F., 1940, *Modern Theory of Solids* (New York and London : McGraw–Hill).
Shockley, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).
Truell, R., and Bronzo, A., 1953, *Phys. Rev.*, **90**, 152.
Wilson, A. H., 1953, *Theory of Metals*, 2nd Edn (Cambridge : University Press).

Properties of Point Contacts on Cobaltite

It has been found by Granville (private communication) that point contacts on single crystals of natural semiconducting cobaltite (CoAsS) give symmetrical voltage–current characteristics. Closely similar peak back-voltages are observed in both directions of current flow. Further experiments on approximately 1 ohm cm cobaltite have been carried out by the author. The contacts under test were made by means of a pointed tungsten wire of 0·005 inch diameter, and each crystal also carried a low resistance contact made by soldering. In the figure the circles show a typical voltage–current characteristic obtained by direct current methods. Different contacts gave peak back-voltages between 5 and 15 volts. The symmetry of the experimental curves shows that this behaviour is not due to a conventional potential barrier at the surface of the crystal. It was suggested that these characteristics arise from the heat dissipation in the vicinity of the point contact and the consequent reduction of spreading resistance. This has been confirmed by examining the voltage–current relations with short (1 μsec) pulses. Heating can be neglected under these conditions. Line A in the figure shows a characteristic of the same contact obtained in this manner. The relation is ohmic within the limits of experimental accuracy and the value of the spreading resistance given by the slope of the line is in reasonable agreement with that estimated from the bulk resistivity of the crystal.

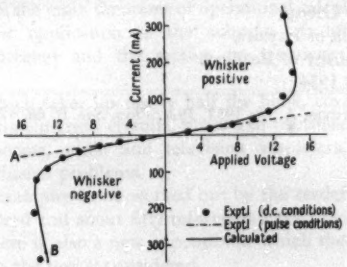


Figure 1. Voltage–current characteristics of a tungsten point contact on cobaltite.

Stukes (1953) has discussed the voltage–current relation to be expected for a small hemispherical contact on a semiconductor in the absence of a barrier, assuming that the resistivity of the bulk material is temperature dependent in accordance with the expression $\rho = \rho_\infty \exp(E/2\boldsymbol{k}T)$. However, for the specimens of cobaltite under investigation the relation between bulk resistivity and temperature could be better represented by an equation of the form $\rho = C(T' - T)$ within the range 20° to 100°C; C and T' are constants. If Stukes' analysis is modified by the use of this expression, the resulting voltage–current relation is found to be

$$V = (C\lambda)^{1/2}(T' - T_0)\left[\frac{x_1 \sinh x_1 + 2(\lambda^*/\lambda)(\cosh x_1 - 1)}{x_1 \cosh x_1 + (\lambda^*/\lambda)\sinh x_1}\right]$$

where $x_1 = (C/\lambda)^{1/2}(I/2\pi r_1)$, V is the voltage applied to the contact, I is the current through the contact, T_0 is the ambient temperature, r_1 is the radius of

the contact, λ and λ^* are the thermal conductivities of cobaltite and whisker respectively. The curve B in the figure is a plot of this equation using $(C\lambda)^{1/2}(T'-T_0)=10{\cdot}3$ volts, $x_1=0{\cdot}023I$ (I in mA) and $\lambda^*/\lambda=3$. Considering the approximations made the agreement with experiment is regarded as satisfactory.

Perhaps a more sensitive test for the correctness of these considerations is provided by an examination of the peak back-voltage and current at various ambient temperatures. The expression in square brackets involving x_1 (i.e. current) is independent of T_0 and voltage turnover corresponds to the condition under which this expression is a maximum. The peak back-current is therefore expected to be independent of ambient temperature and this is substantially in agreement with experimental results. The above equation also implies that the peak back-voltage and the power dissipation at turnover should decrease with increasing ambient temperature and this is likewise in agreement with the observations. In contrast to this Stukes' analysis on the basis of

$$\rho=\rho_\infty \exp(E/2kT)$$

predicts increasing power dissipation at turnover with increasing ambient temperature.

The author wishes to thank Professor R. W. Ditchburn, Dr. H. K. Henisch and the New Zealand Defence Research Organisation for their support.

Semiconductor Group,
Department of Physics,
University of Reading.
12th April 1954.

G. G. E. Low.

Stukes, A. D., 1953, *Proc. Phys. Soc.* B, **66**, 570.

REVIEWS OF BOOKS

Complex Variable Theory and Transform Calculus, by N. W. McLachlan. Pp. xi + 388. 2nd Edn (Cambridge: University Press, 1953). 55*s*.

This is a revised edition of the book published in 1939 under the title of *Complex Variable and Operational Calculus with Technical Applications*, dealing with the use of Laplace transforms and the Mellin inversion theorem in solving certain types of differential equation which arise in technical problems. There are some omissions and additions but, in general, the book follows the first edition fairly closely.

An important feature of the book is the large section devoted to those parts of complex variable theory, namely complex integration, power series expansions, Cauchy's theorem, evaluation of residues and infinite integrals and the Bromwich contour, which are necessary for an understanding of operational methods. Little previous acquaintance with complex variable theory is required of the reader and, although the treatment is more rigorous than in the first edition, it is still suited to those who are not primarily concerned with the mathematical niceties of the subject. One chapter is devoted to the properties of gamma, Bessel and error functions, which are of frequent occurrence in later applications, and another to differentiation and integration under the integral sign.

Part II establishes the main theorems of operational calculus (with illustrative examples), and their application to the solution of differential equations. Impulses are also treated and the section on frequency spectra has been amplified.

The last part, which takes up nearly half the book, covers a wide range of applications, including electrical circuits, filters and transmission lines, vibrating systems, bending beams, radio and television amplifiers, microphones and loudspeakers and diffusion problems.

The number of examples to be worked out by the reader has been increased to more than a hundred and about fifty references have been added to bring the list up to date. There is also a new appendix in which the convergence of the series which occur in the text is considered.

When the first edition was written there were very few textbooks on the practical use of transform theory and it was probably the best. Now that competition is keener, this new edition still maintains its place. A. N. GORDON.

Électrostatique et Magnétostatique, by E. Durand. Pp. xii + 774. (Paris: Masson et Cie, 120 Boulevard Saint-Germain, 1953.) 6,335 fr.

Specialists in theoretical electricity are generally agreed upon the need for a modern treatise on electrostatics and magnetostatics. The reasons may be broadly summarized under three headings: (*a*) the large amount of fallacy and logical inconsistency to be found in the standard treatises, (*b*) the progress made in recent years in the solution of potential problems, (*c*) new phenomena and new fields of application.

The book under review cannot fail to impress the reader by the vast amount of information ordered and systematically expounded in one volume, and no reviewer could attempt to assess the book without some feeling of humility.

Let it be said at once, therefore, that the book does much to fill the need for a new treatise. However, it is necessary to keep in perspective the logical chaos which does exist in the basic theory, and the reader who is conscious of this fact and of the necessity for a complete break from the traditional approach, will find much to criticize. We may give one or two examples. The theory of the electrostatic field for points inside continuous distributions is treated as though, physically at least, it is an obvious extension of the theory for points outside. The difficulties that this leads to when one comes to continuous distributions of polarization are carefully eluded by treating the potential as the basic field variable, but this will not satisfy the connoisseur and, worse still, it will deceive the student. Conductors, rigid dielectrics and electrostrictive dielectrics are, in accordance with the conventional pattern, studied in turn, as though they were physically separable. In fact, one cannot obtain a physically rigorous theory without regarding matter as it really exists—simultaneously chargeable, polarizable and deformable—and in the long run it is simpler to do so. The theory of forces in dielectrics is expounded at great length, but the essence of it is the indirect energy method of Helmholtz, which has been repeatedly and pertinently criticized for the last sixty years. The opening chapters are not, however, without encouraging features. Laplace's equation is developed at the outset and is used where it is customary to use Gauss' induction theorem, so loosely applied (and often loosely proved) in the books we have at present. For this reason alone, Professor Durand's treatment of the principles might be regarded as superior to others, but the fact remains that the approach is, on the whole, the conventional one, and repeats most of the characteristically dubious logic, so that the reader who is looking for a new attitude and a new critical spirit is liable to be disappointed.

Whatever one's feelings on this score may be, there is a stronger case for tolerance than usual with the present book, for nobody interested in the applied mathematical aspects of electrostatics could be unimpressed by the formidable exposition Professor Durand gives us of methods for the solution of potential problems. Each method is developed and applied in a manner clear and concise in style and topical and modern in outlook. To give just one example, some twenty-five closely printed pages are devoted to the discussion and use of the Schwarz transformation, up to such abstruse topics as conductors with bevelled corners. The numerous applications afford the student ample illustration of the points of mathematical technique that arise, and at the same time they are chosen to be of modern practical significance; many electrostatic lens problems are treated, and problems that simulate magnetic ones involving pole-pieces. From the two-dimensional methods, one passes to three-dimensional methods, and, among other topics, practically every system of harmonic potentials is described and applied. Finally, there is an excellent and welcome chapter on numerical methods, where a detailed discussion is given of the finite-difference approximation to Laplace's equation and its solution by the relaxation technique. There are one or two omissions. For example, no mention is made of the application of spherical harmonics to fields with near-spherical boundaries, but it would throw out of perspective the general excellence of this part of the book to expand on such points without also discussing in detail the innumerable merits, which is not possible in a brief review.

Much space is devoted to the discussion of phenomena whose detailed treatment is not customarily regarded as ' electrostatics '. There are complete and

detailed theoretical accounts of piezo-electricity, pyro-electricity, seignetto-electricity and electrets. It is well in keeping with the up-to-date character of the book that seignetto-electricity and electrets should be treated, and this alone fills a gap in the literature. The author does not restrict himself to the framework of macroscopic theory; he treats the various phenomena from the electronic and molecular point of view more or less as completely as is consistent with our established knowledge and with an account so broad in its scope.

About two-fifths of the book is devoted to magnetostatics. Similar general comments may be made regarding the treatment of general principles, applications and new phenomena as with the electrostatics. One or two points, however, deserve special mention. Although, as is usually the case, most of the dubious logic of electrostatics is carried over to magnetostatics, an admirable feature of the present treatment is that the author places the whole of magnetostatics on the basis of the magnetic action of steady currents. This is the correct procedure since we have known for some time that the permanent magnet is the proper magnetic analogue of the electret rather than of electric charge, but to adopt it is an example of the sort of break from tradition that we look for and which it is disappointingly rare to find. It is also gratifying to find that Professor Durand's treatment in other respects does not consist merely of a repetition of what he did for electrostatics, although, of course, some repetition is unavoidable. The extensiveness of the treatment is due to the exposition of para-, dia- and ferromagnetism in the same detail and completeness as was the case with the various dielectric phenomena. Exhaustive accounts are given of practically all the diverse phenomena associated with ferromagnetism, and again, the treatment is modern in its content and outlook, containing most of the significant recent information on the various phenomena, and accounts of such new phenomena as metamagnetism and antiferromagnetism.

Taking the broad view of the book, its most conspicuous characteristics are the great wealth of subject matter and the extraordinary detail and completeness. Almost every topic in physics which has any claim to be called ' electrostatics ' or ' magnetostatics ' is, within reason, treated from every possible angle and to the minutest detail. The author does not hesitate to draw information from any field of theoretical physics or pure mathematics as the need arises. A striking feature is that, everywhere possible, pure theory is applied to specific examples. The text is well illustrated and there are abundant references, both in the text and in comprehensive bibliographies at the ends of most chapters. The book is bound in a hard cover and the production is generally of a good standard. On the whole, the tremendous efforts and pains to which the author must have gone are justified. For all the shortcomings of the book, and whether or not it is destined to be of permanent value as a treatise, the specialist must not miss it, and it should have a place in every university library. R. CADE

Bessel Functions and Formulae, compiled by W. G. BICKLEY. Pp. 12. (Cambridge: University Press, 1953.) 3*s*. 6*d*.

This is a collection of about two hundred formulae reprinted without change from the introduction to volume X of the British Association Tables.

The various kinds of function are defined and all main formulae are listed, e.g. recurrence relations, asymptotic expansions, addition formulae, integral representations, etc. A. N. GORDON.

Structure Reports for 1945–1946, Vol. 10, edited by A. J. C. WILSON. Pp. viii + 325. (Utrecht: Oesthook (for International Union of Crystallography), 1953.) 45 *Dutch fl.* 84*s.*

This volume is one more welcome contribution to filling the gap in recording crystal structures between the last issue of the pre-war *Strukturbericht* and the later volumes of the new *Structure Reports* which are now fast catching up on current researches. The decision was taken by the International Union of Crystallography, at the start of the new venture, to work backwards from that date as well as forwards. This accounts for volume 10 appearing after volumes 11 and 12, in what is apparently the wrong order in time. This publication is of the greatest importance to all those who are concerned with the crystalline state and is an essential part of any serious library of reference. It is also a fine example of book production. D. P. RILEY.

Static Electrification: a Symposium held by the Institute of Physics in London, 25th–27th March, 1953. British Journal of Applied Physics Supplement No. 2. Pp. iv + 104. (London: Institute of Physics, 1953.) 25*s.*

Those who are concerned with static electrification, either from interest or because of its nuisance value, find the literature scattered through a great variety of journals, and going back many years, if, indeed, they find it at all. The symposium held on the subject by the Institute of Physics in March 1953 brought out the importance of its various ramifications and led to valuable discussions arising from different points of view. This is reflected in the papers now published in book form, covering the hazards of static, useful applications, what little is known about the physics of the phenomena, and the design of electrostatic machinery. Turning over the pages of this admirably produced book, one can hardly fail to find something of importance for any immediate problem, and the references to earlier work, though far from complete, form the most useful collection available. W. R. HARPER.

Electrical Breakdown of Gases, by J. M. MEEK and J. D. CRAGGS. Pp. vii + 507. (Oxford: Clarendon Press, 1953.) 60*s.*

In this addition to the publisher's " International Series of Monographs on Physics " the authors have set out to produce a work of reference covering the wide field of breakdown in gases, with particular attention to work published since 1930. They have undoubtedly produced a valuable source book which will find a place in every reference library, together with companion volumes from the same series on the physics of collision processes.

One can sympathize with the authors on the mass of data that is available in what has become a popular subject for research, and in the difficulty of stating firm conclusions at a time when important advances were being made. Since this book will always have to be read with the reservation that it presents the state of the art up to 1950, it is a pity that a brief historical survey could not be included to which the text was related. This would have shown a move, in the 1930's, to abandon the Townsend theory in favour of a ' streamer ' theory of breakdown based upon positive ion space charge and photo-ionization in the gas, but this theory was qualitative only, and the validity of the Townsend theory was still admitted for low values of *pd*. The idea of a transition from one process to

the other was responsible for a considerable amount of post-war research, the results of which began to be published about 1950 and revealed certain deficiencies in the streamer theory, while extending the range of conditions for which the Townsend theory was valid. The true date for the present book is about 1950, but a delay in publication faced the authors with taking account of the latest work. This they have attempted to do by including an extensive appendix divided under headings corresponding to the chapters, but these would have been better located at the end of the individual chapters to which they are a natural continuation.

In the preface the authors state that they have " often used the term streamer to describe the luminous filamentary discharges that are frequently observed to precede the growth of the spark channel, without implications as to their mechanism of formation ". Despite this, the word streamer in the text generally appears to be synonymous with the streamer theory with which one of the authors is identified—in the early chapters on fundamental processes, low pressure discharges, corona, and experimental studies of spark growth references are made to positive and negative streamers and to their junction, and the greater part of the chapter on spark mechanim is devoted to an exposition of the streamer theory as originally developed. Although the authors may now have some doubts concerning the streamer mechanism, this need not condition their quotation of earlier work, but only their comments upon it. It is difficult to appreciate the arguments concerning avalanche to streamer transition in chapter IV on the basis of figures 4.5 and 4.6 (Plate 3), and which are more than a little idealized in the sketch of figure 4.7. In chapter V, on the lightning discharge, it therefore comes as a surprise, on p. 232, that the authors have doubts concerning the discharges from earth as shown on McEachron's photograph, which they might have inserted with advantage. An example of reporting as distinct from commentary is seen in figures 6.9 and 6.10. In the former, formative time lag is shown as a function of gas pressure, whereas in the latter the time lag, under given conditions, appears to be independent of pressure.

Chapter VII, on breakdown characteristics, reviews work leading to an equation $V = Ad + B\sqrt{d}$ for breakdown in a uniform field, but does not refer to the fact that this is not entirely empirical (Whitehead, ref. 51 of this chapter), and this same reference deals with the effects of polarity on asymmetric gaps in advance of reference 64 (p. 310). The important polarity characteristics of the uniform field gap are not mentioned.

A chapter on irradiation and time lags is followed by others on high-frequency breakdown, the spark channel, electrode effects, and glow to arc transitions, much of which is comparatively recent work that has yet to be carried to a conclusion. One does feel that arguments between the theories of incident ions and electron emission at the cathode of arcs are akin to those on the secondary processes in the Townsend discharge, where there may be one or more processes at work according to the circumstances.

The book conforms to the high standard of production for this series and is well indexed, but it would assist the reader if the chapter sub-headings could be given in the table of contents. Some words have been omitted at the foot of p. 179, and an indication of the point and plane electrodes in figures 4.11 and 4.14 would be helpful. Since model testing in the study of lightning protection is mentioned on pp. 233–4, it is regretted that, beyond giving a particular

reference, the authors have not commented upon the validity of such data, of which there is considerable suspicion. In table 11.3 the third figure in the column of sparking voltage should read 32,000 instead of 3,200.

As a factual review this book is a necessary and welcome addition to the literature on the subject, and is likely to remain the standard of reference until another work can finally resolve the Townsend *v.* streamer controversy. Where the authors comment on this they appear to accept limitation and modification to the streamer theory (e.g. wide variation in the ratio of logitudinal to cross field, with little effect on the estimated sparking voltage), and the appendices reveal an ever increasing applicability of the Townsend mechanism. While this is disturbing for the moment, it may well portend the development of a comprehensive general theory, to which we all look forward. F. M. BRUCE.

Analysis of Deformation. Vol. 1. *Mathematical Theory*, by K. SWAINGER. Pp. xix + 285. (London: Chapman and Hall, 1953.) 63*s*.

This well-produced if somewhat expensive volume has evidently been a labour of love to the author, who has sought for a mathematically linear theory of finite deformation and claims that his theory indicates deficiencies in the classical theory of elasticity.

When it is realized that in Swainger's theory (see p. 160) we are led to harmonic problems rather than biharmonic problems and to a constant dilatation in the absence of body force, the extent of the deficiencies claimed by the author is of obvious importance and deserves urgent scrutiny.

In paragraph (11.3), p. 161 *et seq.*, the author is in error in supposing that his equations (11.15) and (11.19) are other than well-known classical results derivable from the fundamental equation of elasticity for the displacement vector $\mathbf{U}$ subject to the author's additional equation of restriction (11.21). This can be written more simply as

$$\operatorname{curl} \mathbf{U} = \text{const}, \qquad \ldots\ldots (1)$$

and it is imperative to discover just how and where this restriction has come into the theory. The reviewer's explanation follows from a consideration of the usual analysis of the infinitesimal deformation of a small volume element at a point P of the strained body into the superposition of a small rigid body displacement of rotation $\tilde{\omega}$ $(=\frac{1}{2}\operatorname{curl}\mathbf{U})$ and of translation $\mathbf{U}$ on a pure straining displacement, the order being immaterial in the superposition of small effects. If the corresponding contributions to the differential displacement $\mathbf{dU}$ at a neighbouring point Q are $\boldsymbol{\delta}\mathbf{U}_1$ due to the rotation and $\boldsymbol{\delta}\mathbf{U}_2$ due to the pure strain, i.e.

$$\mathbf{dU} = \boldsymbol{\delta}\mathbf{U}_1 + \boldsymbol{\delta}\mathbf{U}_2, \qquad \ldots\ldots (2)$$

and we look for the condition that one of these differentials (and therefore the other) is a total differential, we find the condition is precisely the equation (1) above, which is surely significant.

The essence of Swainger's theory is his partition of the differential displacement vector into a whole-body rotation and a straining displacement as

$$\mathbf{dU} = \mathbf{dC}^{\mathrm{wb}} + \mathbf{dD}. \qquad \ldots\ldots (3)$$

Equations (11.4) and (11.5), p. 160, reveal that these two partitions are one and the same. In point of fact the author's equation (10.5) has already imposed the restriction of equation (1) by postulating the existence of the whole-body rotation vector $\mathbf{C}^{\mathrm{wb}}$ and the straining displacement vector $\mathbf{D}$.

All that classical elasticity allows is that

$$\mathbf{dU} = \delta\mathbf{C}^{wb} + \delta\mathbf{D}, \qquad \ldots\ldots (4)$$

and we cannot integrate this to get

$$\mathbf{U} = \mathbf{C}^{wb} + \mathbf{D} + \mathbf{O}, \qquad \ldots\ldots (5)$$

where the constant of integration is a whole-body translation $\mathbf{O}$, and so infer the existence of vectors $\mathbf{C}^{wb}$ and $\mathbf{D}$, unless equation (1) above is satisfied, i.e. unless each element of the body has the same infinitesimal rigid-body rotation. Incidentally it must be reckoned a defect of Swainger's exposition that he never mentions whole-body translation until paragraph (14.4.), p 28, whilst it is tacitly assumed to be zero in equation (1.05) and in figure 1 (10.1), p. 16, and should have appeared as a constant of integration in integrating (11.4) to give (11.5), p. 21.

The ' vector ' $\mathbf{C}^{wb}$ appears first in (10.1) p. 16 as a consequence of the author's fundamental assumption that any finite strain is equivalent to the superposition (in this order, although the author does not say so) of a pure straining displacement ' vector ' $\mathbf{D}$ upon a whole-body rotational or convection displacement ' vector ' $\mathbf{C}^{wb}$. The root cause of all the alleged discrepancies is just this assumption, which the reviewer stigmatizes as being in general invalid for the following reasons.

Would the equivalence be the same if the order of these finite operations were reversed? The author does not consider the addition of finite strains, but we must do so, otherwise the concepts are useless. Would, for example, an alternating sequence of pure strains and finite rotations be equivalent to an alternating sequence of finite rotations and pure strains? This is scarcely likely, since the alteration of order in superposing finite rotations leads to different results. Thus it seems to the reviewer that the only proper passage to the finite strain is by a sequence of incremental strains which imitates more closely the actual straining process. Then *except perhaps in the very special cases where* curl $\mathbf{U}$ *is constant throughout the whole body and remains so at every incremental stage of the whole straining process*, we are not led to the existence of differentiable vectors $\mathbf{C}^{wb}$ and $\mathbf{D}$ at all, and Swainger's theory can therefore only have relevance, at most, to these very special cases, and the deficiencies claimed in the classical theory do not exist. Nor is it necessary to wait for Swainger's second volume with experimental applications and verifications to choose between the theories. Are not the essential truths of Saint Venant's theory of torsion and of flexure with shear well attested by experiment? In both cases curl $\mathbf{U}$ is not constant. Is not the determination of the centre of flexure a well-known experimental study? The very existence of the centre of flexure depends upon curl $\mathbf{U}$ varying across the cross section, and so it could not exist on Swainger's theory. For this and other reasons the reviewer cannot see the theory has any real value.

Exception must also be taken to the section (12), p. 163, where the author embarks on a criticism of the Volterra–Weingarten theory of dislocations without even explaining that a dislocation results from cementing together the suitably prepared edges of a simply connected unstrained body to form a multiply connected strained body, and comes to the conclusion that physically spatial displacement must be continuous and single-valued in a continuous elastic body. This is demonstrably false, the essence of the theory of dislocations being that it does deal with certain real *discontinuities* in $\mathbf{U}$. If we take points P_0, P_0', one on each of the edges which come together at a point P of the

cemented edge or seat of dislocation in the strained body, it follows simply from the definition of **U** that there is a discontinuity in **U** of amount $\overrightarrow{P_0P_0'} \neq 0$, which on the classical theory can be proved to be a rigid body displacement, and is a measure of the existence of a dislocation. If, as Swainger says, **U** " must be continuous ", then there is no dislocation !

The discontinuity gives rise to a cyclic change for those closed circuits which enbrace the hole in the material (see p. 164) and demands a cyclic function such as an angle θ for the description of **U**, but **U** is kept single-valued by restricting the angular range of θ to the physically existing range $0<\theta<2\pi$, the cyclic change in the function on passing once round a curve embracing the hole being related to the cyclic change or discontinuity in **U**. The author seems to think that the use of many-valued functions makes the displacement non-physical, and shows no understanding of the theory he is criticizing, and has only succeeded in postulating real discontinuities out of existence just as he has postulated the unreal vectors $\mathbf{C}^{wb}$ and **D** into existence.

The author's treatment of plane stress and plane strain in section (14), p. 171, starts by over-definition of each of these and falls short of the final simplicities possible with the restriction of equation (1) above. The classical definition of plane stress $\widehat{xz}$ is $=\widehat{yz}=\widehat{zz}=0$; but the author assumes in addition the remaining stresses to be independent of z, whereas this follows from the definition. (Incidentally, in the classical theory free from equation (1), plane stress contains terms in z and z^2.) The classical definition of plane strain is $w=0$, $\partial u/\partial z=\partial v/\partial z=0$. Here we need only take $w=0$; then it follows that $\partial u/\partial z=\partial v/\partial z=0$, and so $\epsilon_{xz}=\epsilon_{yz}=\epsilon_{zz}=0$. The author's assumptions are $\epsilon_{xz}=\epsilon_{yz}=\epsilon_{zz}=0$, with all stresses and strains independent of z. The final simple result, not brought out explicitly by the author, is that this restricted plane strain and plane stress differ only from

$$\widehat{xx}=\frac{\partial^2\chi}{\partial x^2}, \quad \widehat{yy}=\frac{\partial^2\chi}{\partial y^2}, \qquad \widehat{xy}=\frac{\partial^2\chi}{\partial x\,\partial y}, \qquad \nabla_1^2\chi(x,y)=0 \quad \ldots\ldots (6)$$

by the addition of different multiples of the constant dilatation δ to $\widehat{xx}$, $\widehat{yy}$ and $\widehat{zz}$.

On p. 94 Swainger charges Love (*Mathematical Theory of Elasticity*, 4th Edn) with *overlooking* the fact that shear strain should be $\epsilon_{ij}=\frac{1}{2}(\partial U_i/\partial x_j+\partial U_j/\partial x_i)$ for correct invariant transformation with change of coordinate axes. This is just not true. Love was well aware of this (see Love, p. 44). Most mathematical elasticians use the tensor assemblage ϵ_{ij} without apology nowadays. It was pure conservatism which kept Love using his earlier formula, $\epsilon_{ij}=\partial U_i/\partial x_j+\partial U_j/\partial x_i$, as it is with many present-day writers on Applied Elasticity.

The only misprints noted by the reviewer are in the reference, p. 265, and author index, p. 272, where Sir Joseph Larmor appears, almost in Hollywood disguise, as Larmour.

A. C. STEVENSON.

Statistical Methods in Electrical Engineering, by D. A. BELL. Pp. viii + 175. (London : Chapman and Hall, 1953.) 25*s*.

The theory of statistics and probability has long ceased to be the exclusive province of actuaries and biologists, and is now widely applied by physicists and engineers. Radio and communication engineers especially are finding that a knowledge of probability is of very great interest and value to them. The small volume under review is written by an electrical engineer with a wide knowledge of

the literature of his subject, and it deals with such varied topics as curve-fitting, quality control, electrical fluctuations and information theory, to mention but four of the ten chapters. The author has attempted to do something which every reader wants and every writer knows to be difficult, to explain a big subject clearly in a small space. The danger is superficiality, and it is not altogether avoided here. The treatment is in places confused, if not misleading, as for example in chapter four, where the reasons given for taking a product of characteristic functions as a means of deriving the distribution of a difference of errors is quite wrong. In the chapter on information theory it is a great pity to see Shannon's mean power theorem derived once again by a rough and ready method which, though physically appealing, gives no clue to its mathematical precision and generality, especially under the interesting condition when the mean noise power exceeds that of the signal. This is no book for the mathematician, but is not so intended. Apart from questions of mathematical method, however, the many topics are introduced in an easy style which makes for lively reading. The many references have obviously been selected with care and form a valuable feature. The book is very elegantly printed and bound.

P. M. WOODWARD.

CORRIGENDUM

On the Thermal and Electrical Conductivity of Semiconductors, by H. J. GOLDSMID (*Proc. Phys. Soc.* B, 1954, **67**, 360).

It has been pointed out by Dr. R. Berman that the density, ρ, in equation (2) should appear in the numerator. Consequently the values of l_t in table 1 are in error by a factor of ρ^2. The ratio l_t/l_e should be replaced by $\rho^2\, l_t/l_e$ in the last column of table 1 and elsewhere; the conclusion that μ/λ increases with the mean atomic weight remains unaltered.

CONTENTS OF SECTION A

PAGE

Dr. D. H. Parkinson and Mr. J. E. Quarrington. The Molar Heats of Lead Sulphide, Selenide and Telluride in the Temperature Range 20°K to 260°K . 569

Mr. L. S. Kothari. Riesz Potential and the Elimination of Divergences from Meson Theory 580

Miss J. M. B. Lang and Dr. K. J. Le Couteur. Statistics of Nuclear Levels . 586

Mr. K. Ilakovac. Resonant Scattering of Gamma-Rays in ^{63}Cu and ^{56}Fe . . 601

Prof. C. A. Coulson. Note on the Applicability of the Free-Electron Network Model to Metals 608

Dr. J. H. Tait. Two-Group Perturbation Theory in Neutron Transport Theory . 615

Dr. A. R. Bodmer. Spectroscopic Isotope Shift and Nuclear Deformations . 622

Dr. J. Clunie. On Bose–Einstein Functions 632

Research Notes :

Dr. J. C. Barton. Note on the Positive Temperature Coefficient of the Cosmic Radiation 637

Dr. R. P. Rastogi and Mr. R. C. Srivastava. Non-Equilibrium Thermodynamics of Thermal Transpiration of a Dissociating Gas . . . 639

Dr. G. A. Erskine. The Effect of Contaminating Gases on the Energy per Ion Pair in Helium 640

Dr. M. P. Gupta. The Principal Magnetic Susceptibilities of the Maleic Acid Molecule HOOC · HC : CH · COOH 643

Mr. B. Hird and Mr. C. Whitehead. A Modified Scintillation Pair Spectrometer 644

Letters to the Editor :

Mr. F. R. Barclay, Dr. W. Galbraith, Mrs. K. M. Glover, Mr. G. R. Hall and Mr. W. J. Whitehouse. The Spontaneous Fission Rate of ^{240}Pu . 646

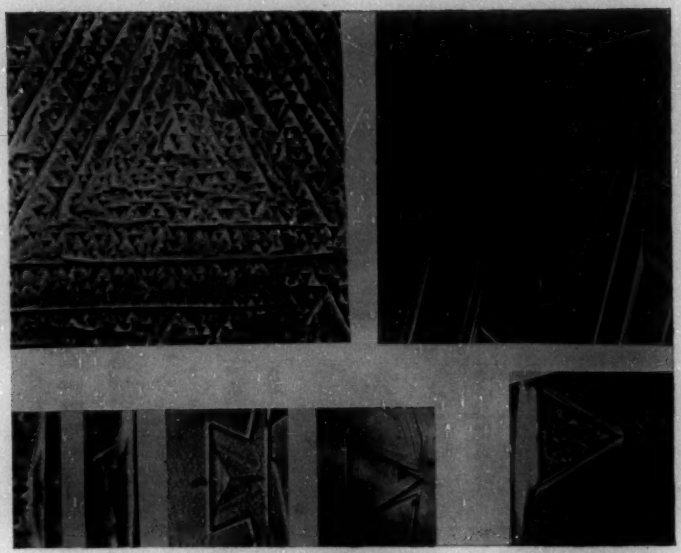
Dr. D. M. S. Bagguley and Dr. N. J. Harrick. The Temperature Dependence of Ferromagnetic Resonance in Colloidal Nickel 648

Reviews of Books 650

Contents of Section B 656

Figure 1.

Figure 2.



(*a*) (*b*) (*c*) (*d*)

Figure 4.

Figure 6.

Figure 1. Part of an octahedral face on a crystallographically very perfect face of a portrait stone. Taken on a silvered (about 90% reflection) crystal face by phase-contrast illumination (×435). The depth of most of the trigons is less than 50 Å.

Figure 2. Deeper trigons (up to a few microns) distributed at random. Note that growth advanced mainly from the right (×80).

Figure 4. Incompleted trigons in different stages of formation.

(*a*) Growth was just stopped along the edge of the tribase (×850).

(*b*) Growth propagates on both sides of the tribase (×850).

(*c*) A more advanced case (×300).

(*d*) A completed trigon, showing more growth-sheets which have not conformed the edges of the trigon (×70).

Figure 6. A large trigon with a flat bottom embedded with many tiny trigons (×300).

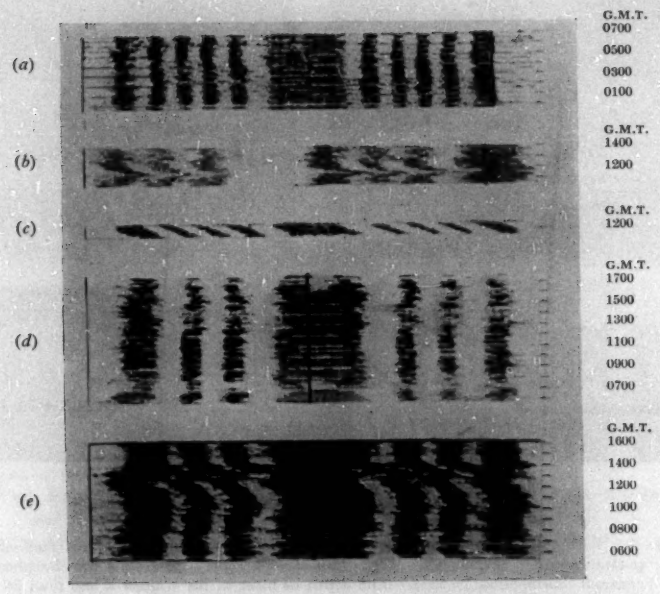
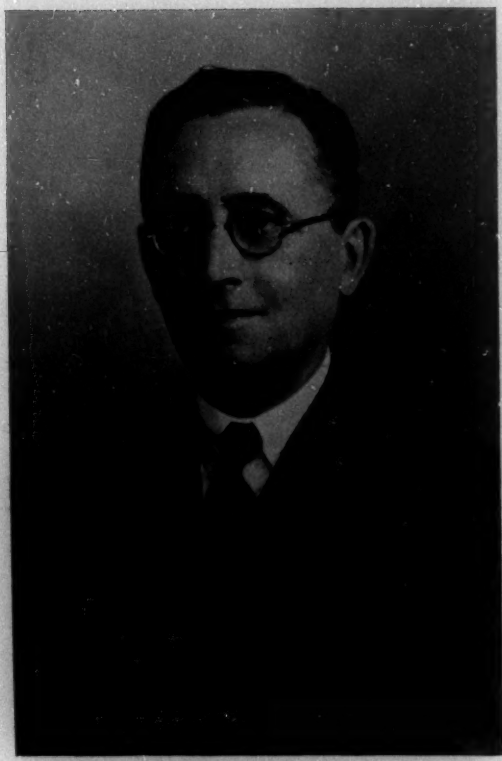


Plate I. Sample records obtained with the rotating radio interferometer.

(*a*) WWVH	15·00 Mc/s	16th July 1953.
(*b*) GRI	9·41 Mc/s	2nd June 1953.
(*c*) —	11·80 Mc/s	7th July 1953.
(*d*) ZQD	9·315 Mc/s	15th July 1953.
(*e*) ZQD	9·315 Mc/s	10th July 1953.

The vertical black line in each section indicates Geographic North.



PROFESSOR R. WHIDDINGTON, C.B.E., M.A., D.Sc., F.R.S.
President of the Physical Society 1952–54.

Physics in the University and the Nation

By R. WHIDDINGTON

Presidential Address, delivered 25th May 1954; MS. received 24th May 1954

Abstract. The importance of physics is to-day so fully realized, and the realization has so widened the sphere of influence of physics, that an insistent demand has arisen for large numbers of well-trained physicists. This has amplified certain difficulties of which the most serious seems to be the shortage of schoolmasters: it is pointed out that, without good teaching in the school, training in the university must take longer and even languish for lack of students. While the number of really first-class physicists passing through the universities has probably for some time remained about the same, the number of well-trained men can perhaps even now under the present threat be increased if steps to this end are taken in time. It is suggested that one step, and one that could be implemented soon, might be to allow and encourage some of the best young physicists now fully engaged in industry and Government Departments to give a little of their time to teaching in schools which are able to avail themselves of such help.

PAST Presidential Addresses to the Physical Society have fallen into two main categories—those dealing with the President's own researches or those delving into more general subjects with a physical bearing. What I shall say this evening will be of this second kind.

I shall ask your indulgence when I occasionally draw on personal recollection and experience, but I shall not allow myself to become too unduly reminiscent.

It was almost fifty years ago that I sat for an Open Scholarship in Natural Science at St. John's College, Cambridge (incidentally using a goose-quill pen, for the first and last time, in answering those deceptively simple looking questions!), so that I can claim, with other contemporaries of course, to have seen many interesting happenings in the world of physics research, teaching and influence, so that I am sure of forgiveness if I am unable to resist an occasional backward glance. This last half-century, not only full of promise but astoundingly fruitful, has understandably attracted numerous competent surveying writers in the various fields of physics; without such résumés and digests those who run could hardly have read! I do not propose to enter this evening any of these fields, but there are some points of difference and change from past and present in the training grounds of physicists—I mean the schools and universities in particular—which are important and worth more than a passing thought. Here I yield to the temptation of a personal recollection in illustration of one point. I took my first degree on Part 2 of the Natural Sciences Tripos (Physics) in 1908, having successfully withstood the great attractions and promise of the biological side, and was determined like every young man, before and since, animated by the spirit of the subject to do some research—in my case at the Cavendish Laboratory. My own experience was, I am sure, typical and representative of the period. Professor J. J. Thomson rejoiced my heart by accepting me and asked whether there was any problem I would like to tackle;

I intimated that I would prefer in the first instance to be guided by him, and was then advised to begin getting apparatus together for vacuum production as he proposed to suggest an experiment on discharge tubes, a subject in which he himself was at that time particularly interested. " Come and see me again ", he said. The great man, for whom we all had the highest admiration and respect, was always rather elusive and difficult to get at, but at last an appointment was made (no full-time secretaries in those days—direct access and hand-written letters were the usual thing) to see him on August 8th. The date itself is of significance; but at the last moment the appointment was cancelled and I received instead a hand-written halfpenny postcard, which I still treasure, addressed to the Cavendish Laboratory, Free School Lane: here is what it said :—" Dear Mr. Whiddington, I did not get an opportunity of talking to you this afternoon about your experiment and I have to leave Cambridge to-morrow. I think the points you might investigate are: (i) When you have freed the electrode from gas, can it recover the gas by standing for several hours at atmospheric pressure or in ammonia ? (ii) Does it recover it when a discharge is passed trying it (*a*) as cathode, (*b*) as anode ? (iii) Are the canalstrahlen to be seen going through the hole in the cathode when the cathode has been freed from air ? (iv) Does an exhausted cathode behave the same way in hydrogen as in air ? Yours very truly, J. J. Thomson."

The atmosphere of the Cavendish then, as now, was one of milling activity, although in those days there were only some thirty or so researchers.

The apparatus available was not only meagre and primitive but extremely scarce and it was just wonderful what was done with glass, Bank of England sealing wax, string and cigarette tins! It was essential to take lessons in glass-blowing (the Professor's research assistant, a past-master in the art of laboratory glass-working, would always oblige for a small fee) since without such self-help and self-reliance nothing ever happened. Discharge tubes had to be made, electrodes fashioned and sealed in, Toepler pump and McLeod gauge had to be conjured out of soft soda glass. The mercury was the rub, but continuous corner searching in summer was rewarding, I recollect, and I found enough and a little to spare. Such was the seamier side of the laboratory and I imagine, too, of other similar penurious institutions in the country. Students, once started, were left very largely to find their own levels, but the spirit of camaraderie was so strong and at all times so helpful, particularly to the newcomer, that the laboratory produced a very satisfactory number of hard-trained men with much pioneer work to their credit. It was the day of self-reliant and energetic individualism, a fact realized to the full by that shining and famous exemplar of these sterling qualities, who was the main attraction to his laboratory in Free School Lane.

Conditions in physics laboratories are very different to-day in many respects —in some ways, many ways, much for the better. There is more apparatus, more workshop, technical and secretarial assistance and certainly more lavish maintenance for the research worker himself. (I subsisted one awkward year on only £60, which, even multiplied by the appropriate money depreciation factor, would still look quite impossible for to-day.) There are some directions however, in which progress may have gone a little too far. A young man, particularly in research, as indeed in life, learns by making mistakes. To-day there is more than a tendency towards pulling him up before he makes them :

he is a member, probably, in one of our great laboratories, of a well-organized team with colleagues each with "an assigned job"; he has a Supervisor and there is a Director rather further in the background. He is well organized. He may be fortunate enough to write a paper either alone or in collaboration—even at this stage he finds he is still in a mistake-preventing machine; keen-minded and lynx-eyed referees scan his manuscript, sometimes refusing it, sometimes passing it, but almost always making suggestions (no doubt excellent ones) for improvement or error elimination. In some ways this supervision is desirable, but blunders are so rarely made on the printed page by young authors that one wonders, sometimes, how many names should really appear on the title-page.

Our best young men are, nowadays, intellectually as good as ever, with the pace and competition getting ever hotter and hotter in our larger schools and much larger universities. It is a good thing, I am sure, for universities to be on the large side, for it is surely true, in general, that for a university to be great it must also be big, although the precise number of undergraduates must, naturally, in any given case, depend on circumstances. In this connection a most striking and interesting phenomenon which has accompanied the rapid growth of the universities of this country during the last three or four decades is the growth in stature of the younger provincial universities: size and status have grown together. May I be allowed to draw on my own experience in illustration: I was privileged to serve the University of Leeds for over thirty years, commencing at a date only fifteen years after the granting of its Charter; it was a most satisfying thing to watch the growth of my Department, proportionately with others, from initially small beginnings to a reasonable size, well staffed and as well equipped as conditions permitted; and it was particularly noteworthy that an increasingly enthusiastic County population were supporting their University not only with money but also with sons and daughters.

I was quite surprised to realize on my retirement that as many as six of my old students had become Professors in universities of high repute (not all in physics incidentally), as also six of my colleagues of the Department, twelve in all. This kind of thing is but one illustration of the increase in stature of the younger universities since university professors but a few decades ago nearly all came from the older universities: the younger universities have definitely grown up! Their success is largely due to the fact that amongst their student intake has always been found a fair proportion of good young men, potentially large academic fish, as it were, who for one reason or another have not been caught in the wide-cast nets of the older universities.

The recent rapid post-war expansion of university undergraduate populations has not, I believe, drawn in many more really first class men; there does not seem to be, as some had thought, an unlimited reservoir of high grade brains of the academic type on which the nation can draw! In my own post-war examining experience (and I have examined in every university in the country except London) the actual number of first-class candidates is little greater than it was. I am not thinking of sheer knowledge—the average candidate for an Honours degree has a quite fantastic volume of knowledge—but I have in mind that intellectual grasp and understanding which any university teacher of experience would immediately recognize and encourage to the utmost. Such boys as these *need* but little teaching or encouragement, they absorb and

understand in the light of their own natures (though more quickly, of course, under slight supervision) and are the cream which one would expect to rise naturally to the top in all examinations. It is the *average* students who require close supervision and coaching rather than suggestion and leading—the process which at schools is often called ' spoon-feeding ' but which at the universities is referred to more kindly as ' coaching '. No one would, at any rate these days, decry the great importance and value of such well-trained young men in our modern world; they are essential and necessary to our acutely scientifically conscious industries and to our rapidly expanded and still expanding Government Research Establishments. The supply, however, is threatened and it is not very difficult to see some of the reasons for the threat, which is a very real and serious one. The basis of the danger is the shortage of schoolmasters, particularly physics masters. The trouble has been brewing for many years in all subjects of the teaching profession, though it is worse in Science than in Arts, and threatens to be quite catastrophic in Physics, largely, I imagine, because of the rewarding and exciting new vistas opened up by beckoning Industry and the Government Scientific Services. When I was a young man, schoolmastering was actually one of the few occupations open to physicists—the possibilities of physics in industry, where chemistry had already won its spurs, were only just beginning to be realized. In those days too, the idea of training science teachers was by no means so firmly entrenched as is it now—I would say in passing that I am of opinion that many possible candidates for specialist school teaching have been rather put off by the necessity of going through a course of training which in prospect they regard as not only over-long but also singularly unattractive and unexciting; in earlier days it was a lot easier to become a physics master, and many then recruited, in what must now seem to the educationists a happy-go-lucky way, were absolutely first class, enthusiastic and successful masters. I was myself fortunate enough to be taught by such an one in my tenderer years and have never forgotten the vigour, enthusiasm and inspiration of his teaching and leadership. During the first few years of my long stay in the young University of Leeds I was struck by the very large proportion of young men in the Honours School of Physics destined to become teachers; it was in the mid-twenties that this bad balance began to be rectified by the start of a small trickle of first class men with Ph.D. degrees into industry (one large concern took five of these men in a few years, and I was to learn later, much to my gratification, that they became known as the ' BIG FIVE ', having risen quickly to high administrative positions). This process was, of course, going on in more or less degree in all universities; more and more scientists—including physicists—entered industry, and with such men in influential directing positions all over the country, industry's attraction for young men was not only enhanced but came to stay. The same is true of our Government Research Establishments, whose reputations, with their sizes, have grown apace.

The result of all this is that the pendulum, so to speak, has swung too far in the direction opposite to that of thirty years ago : to pursue the analogy, we now have too much kinetic energy in industry and not enough in the schools. This is neither the time nor the place to discuss in close detail the reasons for the remarkable change to which I have just directed your attention. I am for myself inclined to think that amongst young men there was never (apart from those admirable few who were ' called ') a general enthusiasm for the teaching

profession, those entering it very frequently regarding it more as a livelihood than a calling; when, added to this, there arose a strong demand for young scientists in well-paid positions in industry, made still more attractive by the prestige attached to the (admittedly overworked) word 'RESEARCH', it is not hard to understand the underlying basis of the present position. I do not believe that money alone is the tap-root of this growth, though it is obviously of much importance: I do suggest, however, that the status of the schoolmaster, high as it is, must be raised still higher by every means in an effort to improve the numbers.

From the viewpoint of the supply of university trained physicists I feel that the real danger point lies in the probability, nay certainty (unless the present trend can be reversed or neutralized in some way), that average boys coming up to the university will have a much lower standard of knowledge; their knowledge will be on a rapid downgrade at a time when university lecture standards will be continuing on a rapid upgrade—in gear with the expansion of physical knowledge and understanding.

I have in my own time seen a 'normal four-year course', leading to what is often now called the 'Special Honours Degree' in physics, reduced to a 'three-year course'. Such reduction in time spent in the university has only been possible because of the higher degree of specialization in the schools, and the consequent taking over, in effect, of a great part of the older first-year course by them. In passing it might be mentioned that some misgivings were expressed, particularly by the more experienced university lecturers and professors, about the disappearance from university courses of the more elementary and basic parts of physics—parts, moreover, which had been traditionally rich in lecture demonstration and experiment. Many thought that not only would there be a change in emphasis within the subject but that also when the time came for graduates of the régime to return to the schools as teachers, they would find themselves called upon to teach those very parts of their subject in which they had received but little reinforcement in the university. The fact stands however that some universities have abandoned their old first-year courses, leaving this basic work to the schools, who, in their turn, will find it increasingly difficult, if not impossible, to carry out such work unless physics teaching, particularly at the sixth form stage, can be safeguarded and strengthened. The general problem is beginning to be widely recognized. A report has quite recently been issued by the National Advisory Council on the Training and Supply of Teachers entitled "Graduate Teachers of Mathematics and Science", under the Chairmanship of Sir Philip Morris, and in their conclusion, which is based on wider considerations than those I employ this evening, appear the following words: "the shortage of graduate teachers of mathematics and science constitutes a national problem . . . there must be serious long-term effects both on the general quality of education of the nation and on the future supply of scientists and technologists". This weighty pronouncement sums up in a few well-chosen words a grave situation: I have tried to point out—in the case of our own subject of physics—the heart of the matter in regard to the school–university relationship, and I incline to the view that swifter action is desirable than 'long-term effects' would seem to require. It is hardly possible, I suppose, to reintroduce immediately an attractive form of the old 'King's Scholarship' scheme the abandonment of which, no doubt, contributed to the present dearth;

nor is it possible, by a stroke of the pen, to conjure a supply of teachers into existence. We may find it best, therefore, to consider how to stave off a portentously evil day by trying to make fuller use of such supplies as at present exist, and I now propose to draw your attention to those possibilities which seem to me not unreasonable. One suggestion would be to make it as easy as possible, *and* attractive enough, for the older teachers of the subject to continue their work well beyond the normal age of retirement; they, after all, have played their part in helping the universities to operate the three-year courses in place of the old-time four-year ones. This age extension plan may, perhaps, be already in use but it can, in any case, only be a temporary palliative. There is, however, an existing reservoir of physicists which might, if it could be suitably tapped, yield a very well worth while number of possible teachers with the requisite knowledge for the senior classes of schools: I refer to the many highly qualified men using their physical knowledge in industrial and Government employment. There must surely be amongst them a proportion who may feel that, after all, they are not so suited to investigational work as they originally thought and who might welcome a change to the teaching profession if that could be arranged. It would be worth much thought—and cost—to try and evolve a workable scheme for the transference of such men, if suitable and wishful, to more congenial employment. Formidable difficulties would, without doubt, be met. How, for example, to encourage a transfer from a better paid to a worse paid profession? How could the 'untrained teacher' problem be overcome? These are just two examples which could be met, no doubt, if the need for action were deemed imperative. Regarding physics in particular, where I believe the problem to be as pressing as anywhere, I feel sure that swift action is highly desirable, and this can only be initiated by high-level Government direction, implemented by departmental machinery.

This is because so few schools remain in this country free enough, of their own volition, to take what action seems good to them, the vast majority being under State aegis and control, and it is these schools which particularly need help in a struggle against odds which are rapidly lengthening as time goes on. A further possibility has also much in its favour, namely, the engagement of professional physicists as part-time teachers in the upper forms of schools. Again it is not hard to envisage difficulties—it is clear at once, for example, that unless the school is on the doorstep of the factory, as it were, there will be loss of time in travel and unequal geographical distribution of opportunity.

These are minutiae. If there are physicists willing to teach, or coach, small groups of eager and promising boys or girls, let them have the chance. Even a few such teachers would, I believe, produce excellent results. In this connection it may be worth while mentioning that, in general, the universities agree that it is on the whole better if a researcher spends some of his time teaching; it is believed in fact that teaching of basic principles adds to the experience and widens the outlook of the researcher—and, of course, greatly benefits the student.

Is it too much to hope that employers, whether Government or industrial, can be pursuaded that in allowing, nay encouraging, researchers to engage in a modicum of teaching they would be more than likely to reap two rewards, firstly an improvement in quality of research and secondly an insurance against a future shortage of scientific man-power?

Of course trends *may* change, but things seem far more likely to get worse rather than better. One thing is certain—if more science teachers cannot be obtained and if the partial remedies suggested prove impracticable, then there is only one way, as far as I can see, of ensuring an adequate supply of physicists for the national needs; the universities will be obliged to take a step backwards and return to the former four-year course system and face up to the prospect of once again welcoming very large numbers of elementary students. This cannot be done without a drop in average standard in the Science Faculties and, with present building capacities, a reduction in the annual output of graduates.

Larger buildings would be needed to deal with the increased first-year entry and, of course, larger staffs to man them; but here is a central point—such staff would certainly be forthcoming to work in universities (where the joys of research could be had), whereas it would be rather unlikely to be forthcoming for the schools, as present experience shows.

It is very relevant, here, to mention that American and Canadian universities commonly have very large first-year classes of an elementary kind, perhaps partly for the reasons we have been considering. Our transatlantic friends have for much longer than ourselves been calling on their universities for large armies of scientists and technologists, and this method seems to suit their purpose—a method which we, in this country, may be obliged to readopt.

The Polarization of Totally Reflected Light

By R. W. DITCHBURN and G. A. J. ORCHARD*
Physics Department, University of Reading

MS. received 5th February 1954, and in final form 26th March 1954

Abstract. This paper describes the measurements of the state of polarization of light totally reflected at a water–air surface. The results agree with theoretical predictions. It is shown that the polarization of light totally reflected at glass–air surfaces can be significantly altered by mild polishing or etching of the surface. Changes of polarization due to the presence of a film of long-chain molecules on a water surface have been observed, and it is suggested that these changes may be used for investigating the properties of liquid films. Some improvements in a photoelectric method of measuring ellipticity are described.

§ 1. Introduction

The state of polarization of light reflected by a surface separating two transparent media may be calculated provided certain boundary conditions are accepted. In this way the well known 'Fresnel relations' were calculated from elastic solid theory and identical equations were later obtained from electromagnetic theory. The essential assumption is that the surface is sharp, i.e. there is a discontinuity in the velocity of propagation though certain components of the vectors are continuous across the boundary. Several workers have investigated the polarization of light reflected in air, or *in vacuo*, at the surfaces of liquids or solids, especially for angles of incidence near the polarizing angle (Jamin 1847, Rayleigh 1892, Bouhet 1931, McBain, Bacon and Bruce 1939, Suhner 1949). For clean surfaces it is found that when the incident light is plane polarized, the reflected light is elliptically polarized instead of being plane polarized. This discrepancy is attributed to a transition layer between the two media. The optical properties of this layer are indeterminate but Drude (1891) has shown that the thickness necessary to satisfy a given set of experimental data has a minimum value if the index of the film is the geometric mean of the indices of the two media. For a clean water surface the residual ellipticity at the polarizing angle is only 0·00042 and the corresponding minimum thickness of the transition later is about 3 Å. The much larger ellipticities produced by monolayers of long-chain molecules may be used to study the properties of these layers.

Measurements on the polarization of light *totally* reflected at the surfaces of glass and quartz prisms were made by Kynast (1907). According to the Fresnel theory the differential phase change Δ between the components polarized in and perpendicular to the plane of incidence is given by the equation

$$\tan \tfrac{1}{2}\Delta = \cos i(\sin^2 i - n_2^2/n_1^2)^{1/2}/\sin^2 i \qquad \ldots\ldots (1)$$

where n_1 and n_2 are the indices of the two media and i is the angle of incidence.

* Now at the British Rayon Research Association, Barton Dock Road, Urmston, nr. Manchester.

Figure 1 shows that there are very large differences between the theoretical and experimental values. The differences are far too large to be due to a transition layer, and they are of the wrong sign. Subsequent work by Volke (1910) showed that the main discrepancy was due to double refraction resulting from strain in the prism. He was able to obtain the theoretical values for Δ by applying small pressures to various faces of the prisms. An agreement produced in this way is not very important because any desired state of polarization of the emergent light could be produced by suitable pressures. Moreover, the predicted variation with wavelength was not obtained.

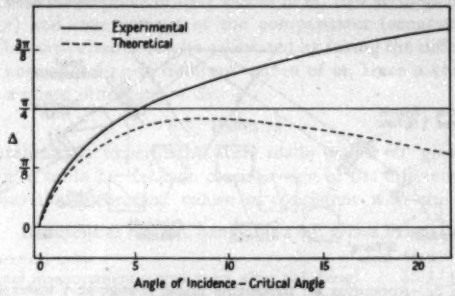


Figure 1. Differential phase change for total reflection at a quartz–air surface. Experimental results due to Kynast ———; theoretical prediction (equation (1)) - - - - - -

In most of the work here reported we used a water surface to avoid the effect of strains in the medium. Four brass tanks with glass windows waxed in position were used. The angles were measured with a spectrometer in a preliminary experiment and corresponded to incidence at 48° 45′, 60° 06′, 69° 44′ and 79° 26′ respectively. The smallest angle is near the critical angle, which is 48° 37′ at 15°C. The tanks were divided into three compartments by walls parallel to the plane of incidence. The water entered the first section, overflowed through small holes in the partitions into the second section (where the light is reflected) and out through the third section. A flow of 5 g min^{-1} was sufficient to maintain a clean surface without causing ripples. The tank was connected through anti-vibration connections to a heavy steel framework which stood on rubber on a concrete block. More rubber was placed between the block and the floor. This arrangement was very effective in eliminating vibration of the water surface.

§ 2. Experimental Method

The totally reflected beam of light is of high intensity and the phase difference between the components is comparable with that obtained in metal optics. It is not necessary to measure the ratio of the amplitudes of the two components but only the differential phase change. This was done with a modified form of the photoelectric method due to Kent and Lawson (1937). The apparatus is shown in figure 2. Light from a sodium lamp is passed through a yellow filter to remove all trace of blue light. After passing the stop it is accurately collimated by the lens and passes through the polarizing prism and also a Soleil–Babinet compensator. It is then totally reflected at the water surface and passes through a piece of

polaroid which is rotating with angular velocity ω. It subsequently enters a photoelectric multiplier cell which is connected to an amplifier tuned to an angular frequency 2ω. The output from the photocell is proportional to $a^2+b^2+(a^2-b^2)\cos 2\omega t$, where a and b are the major and minor axes of the light ellipse. To make a measurement the compensator is adjusted so that after reflection the beam is circularly polarized. The light transmitted by the rotating analyser should then have no component of frequency 2ω and there should be no signal from the amplifier. The phase change introduced by the reflection can then be calculated from the reading of the compensator corresponding to zero signal.

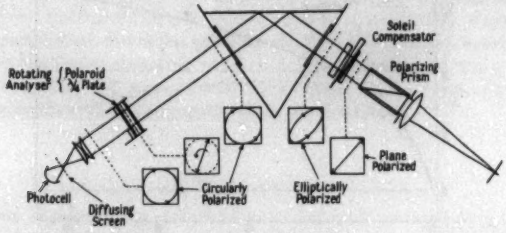


Figure 2. Apparatus for measuring phase change at a water surface.

The accuracy of the measurement is limited by various stray effects. These were investigated by displaying the signal on a cathode-ray screen so as to find the amplitude, frequency and phase of small residual currents.* The most important stray signals are due to (*a*) the dependence of the sensitivity of the photocell on the plane of polarization of the incident light, (*b*) the production, by the rotating unit of a periodic displacement of the position of the beam falling on the photocell cathode (which is not of uniform sensitivity), (*c*) stray light (which is usually partially polarized) passing through the rotating analyser into the photocell. Effects (*a*) and (*b*) were substantially eliminated by attaching a quarter-wave plate to the analyser so that circularly polarized light emerges from the combination and by placing a slightly diffusing screen in front of the photocell. Effect (*c*) was eliminated by suitable light shields. It was also necessary to insert a diaphragm to eliminate beams of light due to internal reflections in the compensator etc.

For all observations the polarizer was set in an azimuth of $\pi/4$ with respect to the plane of incidence. Since the reflection coefficients for the components of the light parallel and perpendicular to the plane of incidence are equal, it was necessary to set the compensator with its axes parallel and perpendicular to the plane of incidence. The phase change α produced by the compensator is then to be added (algebraically) to the differential phase change Δ produced by reflection. We have circularly polarized light when

$$\alpha \pm \Delta = (2m+1)\pi/2 \qquad \ldots\ldots (2)$$

where m is an integer (positive, negative or zero).

* It was subsequently found that J. F. Archard, working at Southampton, had carried out a similar investigation of the accuracy of the Kent and Lawson method. Details of this work are given in his Degree thesis (London University, 1950), and indicated briefly in a recent publication by Archard, Clegg and Taylor (1952)).

The compensator (which will be described elsewhere) gave phase differences between $-\pi$ and 3π for the wavelength of the light used, so values of α corresponding to at least four values of m could be obtained. If α_1 is the value of α which satisfies equation (2) for any given value of m when the plus sign is taken on the left-hand side and α_2 is the corresponding value for the minus sign, then

$$2\Delta = \alpha_2 - \alpha_1. \qquad \ldots\ldots (3)$$

Δ was thus obtained as the difference between a pair of readings of the compensator drum. In order to eliminate various errors Δ was obtained as the mean of 16 such differences. These correspond to four values of m, two settings of the polarizer ($+\frac{1}{4}\pi$ and $-\frac{1}{4}\pi$) and two settings of the compensator (separated by rotation through π). The compensator can be calibrated by taking the difference between drum readings corresponding to different values of m, since a change of 1 in m corresponds to a phase difference of 2π.

§ 3. Results

(*a*) Some preliminary experiments were made with a 60° glass prism. The results are given in table 1. In each case the sign of the difference between the experimental and the theoretical values is consistent with the presence of a

Table 1. Differential Phase Change for a 60° Glass Prism ($n = 1{\cdot}747$)

Theoretical value	46·83°
Original measurement (some years after polishing)	35·3 ± 0·3°
,, ,, (another part of the surface)	37·4 ± 0·3°
After cleaning in alcohol	38·3 ± 0·3°
After immersion in HNO_3 for five minutes	19·0 ± 0·3°
Repeat with another part of the surface	18·8 ± 0·3°
After repolishing	42·05 ± 0·1°
Repeat with another part of the surface	42·55 ± 0·1°

transition layer. For the repolished surface the magnitude of the difference corresponds to a layer about 100 Å thick. It was found that lightly etching the surface, repolishing, etc., produced significant variations. Some variation occurred when the prism was stored for a period of a few days. These results support the view that no accurate test of the theory can be made with solid media. They also show that if a rhomb is to be used for producing circularly polarized light (as suggested by Fresnel) the angles cannot be calculated accurately from the index of refraction but must be adjusted by trial and error. With ordinary glasses and ordinary storage the rhomb will not continue to produce the same phase difference over a long period.

(*b*) The results obtained with pure water (reduced to 15°C) are shown in figure 3. They agree closely with the theoretical curve for 15°C. A detailed analysis of

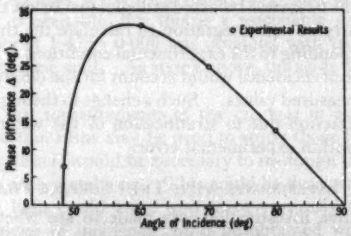


Figure 3. Differential phase change at a water–air surface.

the sources of error was made to see whether there is any significant difference between theory and experiment. The results for different temperatures and conditions are shown in table 2. For angles of incidence of approximately 60°,

Table 2. Measurements with Water

Angle of incidence	Temperature range (°c)	Observed differential phase change	Window correction	Corrected differential phase change	Theoretical value of Δ
48° 45′	10·2–11·3	7·52° ± 0·1°	−0·49°	7·03°	6·92°
48° 45′	20·2–21·0	7·23° ± 0·1°	−0·49°	6·74°	6·18°
48° 45′	34·5–35·3	5·78° ± 0·1°	−0·49°	5·29°	4·49°
60° 06′	18·7–19·0	32·55° ± 0·1°			32·13°
69° 44′	15·3–15·7	24·96° ± 0·1°			25·00°
79° 26′	15·2–15·4	13·98° ± 0·1°	+0·12°	14·10°	13·75°

70° and 80° the measured values exceed the theoretical values by +0·42°, −0·04° and +0·35° respectively. The observational error (estimated from the scatter of results) is ±0·1°. Two possible sources of systematic error were investigated. Firstly, it was found that if the surface of the tank was uncovered (but the water still flowing) the phase difference rose by 0·2°. This was presumably due to contamination of the surface, possibly with dust particles which scatter light. If the cover does not completely eliminate this effect, some positive error will remain. Secondly, the effect of strains in the windows was investigated. The glass used was originally selected to be free from strain by examination between crossed polaroids. It was fixed on with a soft wax (W40 Apiezon), and small strains due to inequalities of temperature should show up as irregularities in the readings. In the earlier measurements with angles of incidence of 60° and 70° the presence of strains was not suspected and, from tests carried out between crossed polarizers, after mounting the glass, appeared to be non-existent. For the other two tanks measurements made with the windows in position showed residual strains producing corrections of −0·5° and +0·1° respectively. Some error must remain after making this correction, and it appears reasonable to assume that the overall error is of order ±0·3°, of which ±0·1° is contributed by observational errors and 0·2° is a systematic error which may tend to have one sign in all the experiments. We may therefore say that theory and experiment agree within 0·3°.

The smallest angle of incidence is very near to the critical angle, and in this region the theoretical differential phase change varies rapidly with angle of incidence and also with temperature because the index varies with temperature. A correction was made for the finite divergence of the beam. Measurements were made at three temperatures. The phase change exceeded the theoretical value by +0·09°, +0·5° and +0·8° at temperatures of 10°, 20° and 35°c. At 35° the critical angle is 48° 41′, and part of the beam was reflected at angles just below the critical angle. It was necessary to carry out an integration to calculate the theoretical differential phase change corresponding to the experimental conditions. A very small change in the assumed angle of incidence would account for the difference of 0·8° between the theoretical and measured values. Such a change in the angle of incidence could be produced by refraction due to stratification of the water. Again the result agrees with theory within experimental error.

§ 4. Experiments with Thin Surface Films

A semi-quantitative investigation was made to see whether changes in the polarization of totally reflected light can be used to investigate monomolecular

films placed on the surface of the liquid. The purpose of the experiments was to see whether the effect of such films was detectable. Hence no special precautions were taken to obtain very pure substances or to avoid contamination by metal ions. The presence of metal ions renders it almost certain that salts of the substances are present in the films (Wolstenholme and Schulman 1950).

Small drops of solutions (in petroleum ether) of certain organic acids were added to the water. These spread on the surface and the petroleum ether evaporated, leaving monomolecular films. A Langmuir trough was used to compress the films.

Figure 4 shows the results obtained for palmitic acid with the 70° tank and with the 48° 45′ tank. The changes in Δ obtained with the 70° tank are detectable,

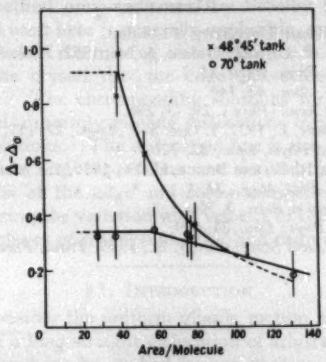


Figure 4. Additional differential phase changes produced by films of palmitic acid. The vertical lines represent limits of error. Additional data (not plotted) show that saturation is obtained for films of surface area less than 35 Å².

but the experimental error is a large fraction of the observed effect. The results obtained with the 48° 45′ tank show a much larger effect (as would be expected on theoretical grounds). Additional data for film areas less than 35 Å² per molecule show that the effect saturates, the increase in the phase change being $0{\cdot}9° \pm 0{\cdot}1°$. Assuming that the film is isotropic and that at this stage the thickness of the film is equal to the length of the molecule found by x-rays, the refractive index can be estimated from the formula given by Drude (1891) and is $1{\cdot}507 \pm 0{\cdot}004$. No direct measurement is available, but this is a reasonable value for this type of substance. An index of $1{\cdot}475 \pm 0{\cdot}006$ was obtained with lauric acid and one of $1{\cdot}528 \pm 0{\cdot}005$ with myristic acid. As stated above, the acids may be contaminated with metallic ions.

It thus appears that measurements of the changes in polarization of totally reflected light due to thin films may be used to study those films. To carry out an extended investigation it would be necessary to re-design the apparatus so as to avoid contamination with metal ions. This could be done with an all glass (or all quartz) tank. The light could enter and leave vertically through the surface and the required angles of incidence could be obtained with suitably inclined reflectors immersed in the water. The surfaces of the reflectors would first be

coated with metal and then with an evaporated layer of SiO_2 or SiO so that the liquid would not be contaminated by metal ions. The reflectors would introduce unknown phase differences but these would not cause an error provided they were constant. With suitable arrangements the measurements on totally reflected light should be about as sensitive as existing polarimetric measurements for surface films.

Acknowledgments

The thanks of one of us (G.A.J.O.) are due to the British Rayon Research Association for the award of a Research Scholarship (during the years 1948–50) and to the Department of Scientific and Industrial Research for a maintenance grant (1947–48).

References

Archard, J. F., 1950, *Thesis*, University of London.
Archard, J. F., Clegg, P. L., and Taylor, A. M., 1952, *Proc. Phys. Soc.* B, **65**, 758.
Bouhet, C., 1931, *Ann. Phys., Paris*, **15**, 5.
Drude, P., 1891, *Wied. Ann.*, **43**, 146.
Jamin, J., 1847, *Ann. Chem.* (*Phys.*), **19**, (3) 296.
Kent, C. V., and Lawson, J., 1937, *J. Opt. Soc. Amer.*, **27**, 117,
Kynast, R., 1907, *Ann. Phys., Lpz.*, **22**, 726.
McBain, J. W., Bacon, R. C., and Bruce, H. D., 1939, *J. Chem. Phys.*, **7**, 818.
Rayleigh, Lord, 1892, *Phil. Mag.*, **33**, 1.
Suhner, F., 1949, *Verres et Refr.*, **3**, No. 1, 3.
Volke, M., 1910, *Ann. Phys., Lpz.*, **31**, 609, .
Wolstenholme, G. A., and Schulman, J. H., 1950, *Trans. Faraday Soc.*, **46**, 475.

Uniformly Moving Dislocations in Anisotropic Media

By R. BULLOUGH† and B. A. BILBY‡
Department of Metallurgy, University of Sheffield

Communicated by A. G. Quarrell; MS. received 23rd April 1954

Abstract. Solutions which correspond to surface waves are given of the general dynamic equations of anisotropic elasticity when all displacements are independent of one coordinate. These solutions are used to discuss the uniform glissile motion of a long straight dislocation line of arbitrary Burgers vector by a generalization of a method previously used by Eshelby for the isotropic case. Although the method used here is generally applicable, explicit solutions are for simplicity given only when the moving dislocation line is so related to the symmetry elements of the crystal that the edge and screw components can be considered separately. The corresponding solutions for static dislocations in anisotropic media and uniformly moving dislocations in isotropic media can be derived from these solutions. The dislocation line is also treated in the Peierls–Nabarro approximation and expressions are derived for the variation with velocity of the widths of the edge and screw components. The results are illustrated by considering the variation with velocity of the widths of dislocations moving in the basal planes of magnesium, zinc and cadmium.

§ 1. Introduction

In this paper we consider the uniform glissile motion through an anisotropic elastic medium of a long straight dislocation of arbitrary Burgers vector.

The method used can be applied for any position of the dislocation line, but as the analysis is clumsy if the position of the line is quite unrestricted, explicit solutions are given only when the moving dislocation line is so related to the symmetry elements of the crystal that the edge and screw components can be considered separately.

In § 2 we derive the fundamental solutions for the required surface waves. These are applied in § 3 and § 4 to obtain expressions for the displacements round the moving edge and screw dislocations. In § 5 we show how results previously obtained for the moving dislocations in an isotropic medium (Eshelby 1949 a) and for static dislocations in an anisotropic medium (Eshelby 1949 b, Eshelby, Read and Shockley 1953) follow from these expressions. The extension of the results to the Peierls–Nabarro approximation is discussed in § 6 and expressions for the variation of dislocation width with velocity are given. The results are applied in § 7 to find the widths of dislocations in magnesium, zinc and cadmium.

§ 2. Surface Waves in Anisotropic Elastic Media

The general dynamic equations for an anisotropic elastic medium are:

$$c_{ijkl}\frac{\partial^2 u_k}{\partial x_j \partial x_l} = \rho \frac{\partial^2 u_i}{\partial t^2}. \qquad \ldots\ldots(1)$$

† C. G. Carlisle Scholar, Department of Metallurgy, The University, Sheffield.

‡ Royal Society Sorby Research Fellow and Honorary Lecturer in the Department of Metallurgy, The University, Sheffield.

Here u_k is the displacement referred to Cartesian coordinates x_i, ρ is the density, and c_{ijkl} the elastic constant tensor.

We take the dislocation line parallel to x_3 and moving in the x_1 direction with velocity c. Our method of solution is similar to that of Eshelby (1949 a). We imagine the material cut along the x_1x_3 plane, and apply suitable moving tractions to the cut surfaces, so that on rewelding we obtain a moving dislocation which is not acted upon by any external forces or couples. Thus we require solutions of (1) corresponding to surface disturbances propagating in the $x_2=0$ plane and vanishing as $|x_2|\to\infty$. The solution for a long straight dislocation lying parallel to the x_3 axis will be independent of the x_3 coordinate. When this is so, a solution of (1) corresponding to surface waves propagating in the x_1 direction is

$$u_j(x_1', x_2)=\sum_n C_n P_{jn}\exp\{+s(-\lambda_n x_2+ix_1')\} \qquad \ldots\ldots(2)$$

where $n=1, 2, 3$, $P_{1n}=S_nD_n+R_nE_n$, $P_{2n}=L_nE_n+R_nD_n$, $P_{3n}=R_n^2-S_nL_n$, and the C_n are arbitrary complex constants. λ_n is a root of the sextic equation

$$\begin{vmatrix} L(z) & -R(z) & D(z) \\ -R(z) & S(z) & E(z) \\ D(z) & E(z) & F(z) \end{vmatrix}=0$$

and $x_1'=x_1-ct$. If (2) is to represent a surface wave, the real part of λ_n must not vanish: this will in general impose conditions on the velocity of the surface waves. The quantities R_n, S_n, L_n, D_n, E_n, F_n are of the form

$$R_n=R(\lambda_n)=h_1\lambda_n^2+ih_2\lambda_n+h_3$$

where the values of h_1, h_2, h_3 are given in the following table:

	h_1	h_2	h_3
S_n	c_{22}	$-2c_{26}$	ρc^2-c_{66}
R_n	$-c_{26}$	$c_{12}+c_{66}$	c_{16}
L_n	c_{66}	$-2c_{26}$	ρc^2-c_{11}
D_n	c_{46}	$-(c_{14}+c_{56})$	$-c_{15}$
E_n	c_{44}	$-2c_{45}$	ρc^2-c_{55}
F_n	c_{24}	$-(c_{25}+c_{46})$	$-c_{56}$

In this table the standard contractions for pairs of indices of the elastic constant tensor are used, namely that 11, 22, 33, 23, 31, 12 are replaced by 1, 2, 3, 4, 5, 6 respectively. The method of solution outlined in § 3 could be applied using the general displacements (2), but the resulting expressions are very clumsy. We shall therefore treat explicitly only the problem arising when

$$c_{14}=c_{15}=c_{24}=c_{25}=c_{34}=c_{35}=c_{46}=c_{56}=0,$$

that is, all c_{pqrs} with an odd number of suffixes equal to 3 vanish. In a monoclinic crystal, for example, this condition is satisfied when the x_1x_2 plane is a symmetry plane or x_3 is a diad axis. The equations (1) then reduce to the two equations

$$c_{\alpha\beta\varepsilon\sigma}\frac{\partial^2 u_\varepsilon}{\partial x_\beta \partial x_\sigma}=\rho\frac{\partial^2 u_\alpha}{\partial t^2} \qquad \ldots\ldots(3)$$

and

$$c_{55}\frac{\partial^2 u_3}{\partial x_1^2}+2c_{45}\frac{\partial^2 u_3}{\partial x_1\partial x_2}+c_{44}\frac{\partial^2 u_3}{\partial x_2^2}=\rho\frac{\partial^2 u_3}{\partial t^2}. \qquad \ldots\ldots(4)$$

In (3) we use the convention that the summation range for Greek subscripts is only over 1, 2. The equations (3) determining u_1 and u_2 on the one hand and that determining u_3 are thus not related. Also the stresses p_{11}, p_{22}, p_{12}, p_{33} depend only on the derivatives of u_1 and u_2 while p_{23}, p_{31} depend only on those of u_3

(Eshelby, Read and Shockley 1953). Thus we can consider the problems with $u_1=u_2=0$, $u_3\neq 0$ and with $u_1\neq 0$, $u_2\neq 0$, $u_3=0$ separately. A suitable solution of (3) is

$$\left.\begin{aligned} u_1&=\psi(A_n, S_n, \lambda_n)\\ u_2&=\psi(A_n, R_n, \lambda_n)\end{aligned}\right\} \qquad \ldots\ldots(5)$$

where

$$\psi(a_n, b_n, c_n)=\sum_n a_n b_n \exp\{s(-c_n x_2+ix_1')\}, \qquad \ldots\ldots(6)$$

and the summation extends over $n=1, 2$. A_1, A_2 are arbitrary complex constants and S_n, R_n are the quantities defined in the table above, namely

$$S_n=S(\lambda_n)=c_{22}\lambda_n^2-c_{66}+\rho c^2-2ic_{26}\lambda_n, \qquad \ldots\ldots(7)$$

$$R_n=R(\lambda_n)=-c_{26}\lambda_n^2+c_{16}+i(c_{12}+c_{66})\lambda_n. \qquad \ldots\ldots(8)$$

λ_1 and λ_2 must be roots of the quartic

$$S(z)[c_{66}z^2-c_{11}+\rho c^2-2ic_{16}z]=[R(z)]^2. \qquad \ldots\ldots(9)$$

By writing $z=iz'$ in (9) we obtain a quartic all of whose coefficients are real, and whose roots are therefore of the form $q_n\mp ip_n$, where $n=1, 2$ and p_n, q_n are real. $p_n\neq 0$ provided that the velocity is suitably restricted. The roots of (9) are thus of the form $\pm p_n+iq_n$; the positive sign is appropriate for a surface disturbance of the half-medium $x_2>0$.

A suitable solution of (4) is

$$u_3=\chi(A,\gamma)=A\exp[s(-\gamma x_2+ix_1')] \qquad \ldots\ldots(10)$$

where A is an arbitrary complex constant, and γ must be a root of the quadratic

$$c_{44}z^2-2ic_{45}z+c^2\rho-c_{55}=0. \qquad \ldots\ldots(11)$$

The roots of this quadratic are of the form $\gamma=\pm a+ib$ where a and b are real constants, $a>0$, provided that $\rho c^2<c_{55}-c_{45}^2/c_{44}$.

§ 3. The Moving Edge Dislocation

In the general anisotropic case here discussed it is not possible to reduce the problem to one in the half-medium $x_2>0$ by considering the symmetry. We must therefore impose moving displacements on both the upper and lower half-media obtained by cutting along $x_2=0$. For the upper half, $x_2>0$, we take:

$$u_1=\mathscr{R}\psi(A_n, S_n, p_n+iq_n), \qquad \ldots\ldots(12)$$

$$u_2=\mathscr{R}\psi(A_n, R_n, p_n+iq_n) \qquad \ldots\ldots(13)$$

and for the lower, $x_2<0$,

$$u_1^*=\mathscr{R}\psi(A_n^*, S_n^*, -p_n+iq_n), \qquad \ldots\ldots(14)$$

$$u_2^*=\mathscr{R}\psi(A_n^*, R_n^*, -p_n+iq_n). \qquad \ldots\ldots(15)$$

Here ψ is given by (6). A_1^*, A_2^* are further arbitrary complex constants, and $S_n^*=S(-p_n+iq_n)$, $R_n^*=R(-p_n+iq_n)$, S and R being given by (7) and (8). It is readily verified that S_n^* and R_n^* are the complex conjugates of S_n and R_n respectively. We can now write down more general displacements U_1, U_2, U_1^*, U_2^* by replacing ψ in (12), (13), (14) and (15) respectively by

$$\Psi=\int_0^\infty f(s)\psi\, ds. \qquad \ldots\ldots(16)$$

Let

$$A_n=A_{nr}+iA_{ni} \qquad \ldots\ldots(17)$$

and similarly for A_n^*, S_n, R_n. Then we have to choose the eight real constants A_{nr}, A_{ni}, A_{nr}^*, A_{ni}^* and the function $f(s)$ so that our boundary conditions are satisfied. These are obtained from the following considerations.

If the motion of the dislocation is to be glissile, the Burgers vector must lie in the plane of motion, except in the case of the transformation dislocation (Bilby 1953). For simplicity we take the Burgers vector to be $(b_0, 0)$; the extension to the more general case is tedious but straightforward. Then we have on the plane $x_2=0$

$$\left.\begin{aligned} U_1 - U_1^* &= -b_0/2 \qquad x_1'>0 \\ U_1 - U_1^* &= +b_0/2 \qquad x_1'<0 \end{aligned}\right\} \qquad \ldots\ldots(18)$$

$$U_2 - U_2^* = 0. \qquad \ldots\ldots(19)$$

The last condition must hold for all x_1' except at $x_1'=0$, where there is a singularity. The stresses $p_{\alpha 2}$, $p_{\alpha 2}^*$, $\alpha=1, 2$ derived from u_α, u_α^* by the relations

$$p_{\alpha\beta} = c_{\alpha\beta\gamma\delta}(\partial u_\gamma/\partial x_\delta) \qquad \ldots\ldots(20)$$

are the Fourier transforms of the stresses $P_{\alpha 2}$, $P_{\alpha 2}^*$ similarly derived from U_α, U_α^*. On $x_2=0$, $p_{\alpha 2}$, $p_{\alpha 2}^*$ are all of the form $C_1 \cos sx_1' + C_2 \sin sx_1'$, where the C_1, C_2 are functions of p_n, q_n and the constants A_{nr}, A_{nr}^* etc. Denoting the cosine and sine terms in these expressions by the superscripts E and O, we write

$$p_{\alpha 2}(x_1', 0) = p^{\mathrm{E}}{}_{\alpha 2} + p^{\mathrm{O}}{}_{\alpha 2} \qquad \ldots\ldots(21)$$

and similarly for $p_{\alpha 2}^*$. Then we require that

$$p^{\mathrm{E}}{}_{\alpha 2} = p^{\mathrm{E}}{}_{\alpha 2}^* = 0 \qquad \text{all } x_1'. \qquad \ldots\ldots(22)$$

This is sufficient to ensure (Eshelby 1949 a, Doetsch 1937) that the resultant forces X_α, X_α^* on the upper and lower surfaces $x_2=0$, are zero: that is,

$$X_\alpha = \int_{-\infty}^{+\infty} P_{\alpha 2}(x_1', 0)\, dx_1' = 0, \qquad \ldots\ldots(23)$$

with a similar equation for X_α^*.

In order that the surfaces may be rewelded without the application of additional forces to the boundary plane, we must also ensure that

$$P_{\alpha 2}(x_1', 0) - P_{\alpha 2}^*(x_1', 0)$$

vanishes. Accordingly, we require that

$$p_{\alpha 2}(x_1', 0) - p_{\alpha 2}^*(x_1', 0) = 0 \qquad \text{all } x_1'. \qquad \ldots\ldots(24)$$

$P_{\alpha 2}(x_1', 0) - P_{\alpha 2}^*(x_1', 0)$ now vanishes 'almost everywhere', and a condition similar to (23) ensures that any singularity at $x_1'=0$ does not correspond to a point force. Finally, we set

$$u_2(x_1', 0) = u_2^*(x_1', 0) \qquad \text{all } x_1'. \qquad \ldots\ldots(25)$$

This means again that (19) is satisfied, except at $x_1'=0$, and that here the singularity is limited by another condition like (23). It is readily verified that, in view of (21) and (22), the conditions (24) and (25) are satisfied by choosing:

$$\left.\begin{aligned} A_{nr} &= A_{nr}^* \qquad n=1,2 \\ A_{ni} &= -A_{ni}^* \qquad n=1,2. \end{aligned}\right\} \qquad \ldots\ldots(26)$$

Inserting (26) in (12)–(15) and setting

$$\sum_n (A_{nr} S_{ni} + A_{ni} S_{nr}) = -1 \qquad \ldots\ldots(27)$$

so that the amplitude of the sinusoidal components is unity, we have finally on $x_2=0$

$$u_1, u_1^* = \sum_n [(A_{nr}S_{nr} - A_{ni}S_{ni})\cos sx_1' \pm \sin sx_1'], \quad \ldots\ldots(28)$$

$$u_2 = u_2^* = \sum_n [(A_{nr}R_{nr} - A_{ni}R_{ni})\cos sx_1']. \quad \ldots\ldots(29)$$

In (28) the negative sign refers to u_1^*. Writing the condition (18) in the form

$$U_1(x_1', 0) - U_1^*(x_1', 0) = -\frac{b_0}{\pi}\int_0^\infty \frac{\sin sx_1'}{s}\,ds \quad \ldots\ldots(30)$$

and comparing with (16) and (28) we have

$$f(s) = -\frac{b_0}{2\pi}\left(\frac{1}{s}\right). \quad \ldots\ldots(31)$$

We have now to insert (31) in (16) and evaluate the displacements $U_\alpha(x_1', x_2)$, $U_\alpha^*(x_1', x_2)$ and the stresses derived from them by (20). Using the integral

$$\int_0^{+\infty} \frac{\exp\{-s(x+iy)\}}{s}\,ds = -\ln(x+iy), \qquad x>0 \quad \ldots\ldots(32)$$

we find, after some reduction, that the displacements both for $x_2>0$ and $x_2<0$ may be written

$$U_1 = \mathscr{R}[(b_0/2\pi)\sum_n A_n S_n \ln(z_n)], \quad \ldots\ldots(33)$$

$$U_2 = \mathscr{R}[(b_0/2\pi)\sum_n A_n R_n \ln(z_n)] \quad \ldots\ldots(34)$$

where

$$z_n = x_1' + i(p_n + iq_n)x_2 \quad \ldots\ldots(35)$$

and the A_n are given as functions of $c_{\alpha\beta\gamma\delta}$, ρ and c by (22), (24), (25) and (27). Using (33) and (34), a simple calculation shows that there is zero resultant couple acting on any circuit containing $x_1'=0$ for arbitrary choice of A_n. This, together with (23), shows that the material is in equilibrium.

§ 4. The Moving Screw Dislocation

The dislocation line of Burgers vector $(0, 0, b_0)$ lies along the x_3 axis and moves with velocity c in the x_1 direction. As in § 3, we cut the x_1x_3 plane and impose moving displacements u_3, u_3^* on the half-media $x_2>0$, $x_2<0$. If the x_1x_2 plane is a symmetry plane, the problem can in fact be reduced to one in the region $x_2>0$, but the resulting simplification is a trivial one in this case, so we follow the same method as in § 3. We take

$$\left.\begin{aligned} u_3(x_1', x_2) &= \mathscr{R}\chi(A, a+ib) \\ u_3^*(x_1', x_2) &= \mathscr{R}\chi(A^*, -a+ib) \end{aligned}\right\} \quad \ldots\ldots(36)$$

where χ is given by (10), and denote by U_3, U_3^* the corresponding generalizations obtained by using (16) with $\psi=\chi$. Then our first boundary condition is

$$\left.\begin{aligned} U_3 - U_3^* &= -b_0/2 \qquad x_1'>0 \\ U_3 - U_3^* &= +b_0/2 \qquad x_1'<0. \end{aligned}\right\} \quad \ldots\ldots(37)$$

Let $p_{\alpha 3}$, $p_{\alpha 3}^*$, $P_{\alpha 3}$, $P_{\alpha 3}^*$ be the stresses derived from u_3, u_3^*, U_3, U_3^* respectively by (20) with $\beta=\gamma=3$, and denote their even parts by the superscript E. Then it is readily verified that the boundary conditions

$$p^{\mathrm{E}}{}_{23}(x_1', 0) = p^{\mathrm{E}}{}_{23}{}^*(x_1', 0) = 0, \quad \ldots\ldots(38)$$

$$p_{23}(x_1', 0) = p_{23}^*(x_1', 0) \quad \ldots\ldots(39)$$

are satisfied if $A_r = A_r^* = 0$ and $A_i = -A_i^*$. Taking $A_i = -1$ we thus obtain

$$u_3, u_3^* = \pm \exp(\mp asx_2) \sin s(x_1' - bx_2). \quad \ldots\ldots (40)$$

The upper sign refers to u_3. Writing (37) in the form (30) and comparing this with the generalization (16) of (40) when $x_2 = 0$, we again find that $f(s)$ is given by (31). By using the equivalent of (16) we can evaluate U_3, U_3^* and the stresses derived from them by using (20) with $\beta = \gamma = 3$. Using the imaginary part of (32), we find that the displacement both for $x_2 > 0$ and $x_2 < 0$ may be written

$$U_3 = -\frac{b_0}{2\pi} \tan^{-1}\left[\frac{x_1' - bx_2}{ax_2}\right] \quad \ldots\ldots (41)$$

where, from (11),

$$a = [c_{44}c_{55} - c_{45}^2 - c_{44}\rho c^2]^{1/2}/c_{44}, \quad b = c_{45}/c_{44}.$$

§ 5. Comparison with the Static and Isotropic Solutions

U_1 and U_2 have been given explicitly for the static edge dislocation when $c_{16} = c_{26} = 0$ (Eshelby 1949 b). This condition is satisfied, for example, if the x_1x_3 plane is a symmetry plane, or if x_2 is a diad axis. When $c_{16} = c_{26} = 0$, the quartic (9) becaomes a quadratic for z^2 with real coefficients. If the roots of this are real and unequal, the roots of the quartic are $\pm p_1$, $\pm p_2$, while if they are complex conjugates $p_1 = p_2 = p$ and $q_1 = -q_2 = q$ and the roots of the quartic are $\pm p \pm iq$.

When the roots of the quadratic are real and unequal, $S_{ni} = R_{nr} = 0$. (21), (22) and (24) then give, in general, $A_{1r} = A_{2r} = 0$, while (22), with $\alpha = 2$, and (27) determine the A_{ni} as functions of c_{11}, c_{12}, c_{22}, c_{66} and ρc^2. (33) and (34) become

$$U_1 = -(b_0/2\pi) \sum_n [A_{ni} S_{nr} \tan^{-1}(p_n x_2/x_1')], \quad \ldots\ldots (42)$$

$$U_2 = -(b_0/2\pi) \sum_n [A_{ni} R_{ni} \ln (x_1'^2 + p_n^2 x_2^2)^{1/2}]. \quad \ldots\ldots (43)$$

In plane strain the relations between the stress components P_1, P_2, P_6, and the strain components E_1, E_2, E_6, may be written

$$P_i = c_{ij} E_j, \quad \ldots\ldots (44)$$

$$E_i = s_{ij}' P_j \quad \ldots\ldots (45)$$

where $i, j = 1, 2, 6$ and s_{ij}' is the matrix reciprocal to c_{ij}. When $c = 0$, the p_n become identical with Eshelby's λ_n, and on expressing the $A_{ni}S_{ni}$ and $A_{ni}R_{ni}$ in terms of the s_{ij}' (42) and (43) reduce to his equations (8). There are some obvious printing errors in his expression for v.

When the roots of the quadratic are complex conjugates, S_2 and iR_2 are the complex conjugates of S_1 and iR_1, respectively. (21), (22) and (24) give in general $A_{1r} = -A_{2r}$, $A_{1i} = A_{2i}$ and from (27) $A_{1r}S_{1i} + A_{1i}S_{1r} = A_{2r}S_{2i} + A_{2i}S_{2r} = -\frac{1}{2}$. (33) and (34) become

$$U_1 = \left(\frac{b_0}{4\pi}\right)\left[\tan^{-1}\left(\frac{px_2}{x_1' - qx_2}\right) + \tan^{-1}\left(\frac{px_2}{x_1' + qx_2}\right) + (A_{1r}S_{1r} - A_{1i}S_{1i}) \ln\left\{\frac{(x_1' - qx_2)^2 + p^2x_2^2}{(x_1' + qx_2)^2 + p^2x_2^2}\right\}\right], \quad \ldots\ldots (46)$$

$$U_2 = \left(\frac{b_0}{4\pi}\right)\left\{(A_{1r}R_{1r} - A_{1i}R_{1i}) \ln [(x_1' - qx_2)^2 + p^2x_2^2][(x_1' + qx_2)^2 + p^2x_2^2] + 2(A_{1r}R_{1i} + A_{1i}R_{1r})\left[\tan^{-1}\left(\frac{px_2}{x_1' + qx_2}\right) - \tan^{-1}\left(\frac{px_2}{x_1' - qx_2}\right)\right]\right\}. \quad \ldots\ldots (47)$$

A_{1l} and A_{1r} are determined as functions of the c_{ij} and ρc^2 by (22) with $\alpha=2$ and (27). Putting $c=0$ and expressing the constants in terms of the s_{ij}', we obtain Eshelby's equations (10).

As in the static case, when the discriminant of the quadratic vanishes, the results for U_1, U_2 can be derived from (42), (43) by a limiting process.

Formulae for the moving isotropic edge dislocation follow from (42), (43). We have $c_{11}=c_{22}=\lambda+2\mu$, $c_{12}=\lambda$, $c_{66}=\mu$ and $p_n^2=1-c^2/c_n^2$, where c_1, c_2 are the velocities of dilatational and shear waves. U_1 and U_2 reduce at once to the expressions already published (Eshelby 1949 a) (eqns. (12), (13)).

The comparison of our results with those published for the anisotropic screw dislocation is immediate, since the latter are given in terms of the c_{ij} (Eshelby, Read and Shockley 1953). On setting $c=0$ in (41), and comparing roots of the two quadratics, we obtain equation (7.6) of their paper. Although it is not stated by Eshelby, Read and Shockley, the imaginary part of D, given by (7.8), is in fact zero, as is obvious from their equation (7.4).

§ 6. The Moving Peierls–Nabarro Dislocations

The calculations have now to be extended to the Peierls–Nabarro dislocation model, to determine the variation of dislocation width with velocity. We can use either the original method (Nabarro 1947) or employ a continuous distribution of elementary Volterra dislocations (Eshelby 1499 b). The latter procedure is used here.

From (33) and (34) we obtain

$$P_{12}(x_1',0)=\frac{b_n}{2\pi}\cdot\frac{K(c)}{x_1'} \qquad \ldots\ldots(48)$$

where

$$K(c)=\mathscr{R}\{\sum_n A_n[c_{16}S_n+c_{66}R_n+i\lambda_n(c_{26}R_n+c_{66}S_n)]\} \qquad \ldots\ldots(49)$$

and A_n are given as functions of $c_{\alpha\beta\gamma\delta}$, ρ and c by (21), (22), (24), (25) and (27). When $c=0$, $K(c)$ reduces to the K of the static analysis (Eshelby 1949 b). We can thus find the energy of the dislocation by using (16E) (i.e. Eshelby 1949 b, eqn. (16)). Repetition of this analysis (eqns. (17E)–(20E)) leads to a relative displacement $\phi=u_1-u_1^*$ given by

$$\phi/b_0=-(1/\pi)\tan^{-1}(x_1'/\xi) \quad \text{with} \quad \xi=\frac{K(c)a_0}{2c_{66}} \qquad \ldots\ldots(50)$$

where b_0 is the Burgers vector and a_0 is the spacing of the atomic planes in the x_2 direction. ξ is the width of the dislocation and is a measure of the distance along the x_1-axis within which the departure from Hooke's law is important.

In the case of the screw dislocation we have from (41)

$$P_{23}(x_1',0)=\frac{b_0}{2\pi}\frac{L(c)}{x_1'} \qquad \ldots\ldots(51)$$

where

$$L(c)=[c_{44}c_{55}-c_{45}^2-c_{44}\rho c^2]^{1/2}. \qquad \ldots\ldots(52)$$

A repetition of the above analysis leads to a relative displacement $\phi=u_3-u_3^*$ given by

$$\frac{\phi}{b_0}=-\frac{1}{\pi}\tan^{-1}\left(\frac{x_1'}{\eta}\right) \quad \text{with} \quad \eta=\frac{L(c)a_0}{2c_{44}}. \qquad \ldots\ldots(53)$$

η is the width of the moving screw dislocation.

§ 7. The Variation of Dislocation Widths with Velocity

The above results will be illustrated by considering the widths of dislocations moving in the basal planes of the hexagonal close-packed metals magnesium, zinc and cadmium. If the axes of § 3 are taken so that x_2 is along [0001] and x_1 and x_3 are any two orthogonal directions lying in (0001), we have always $c_{14}=c_{15}=c_{24}=c_{25}=c_{34}=c_{35}=c_{45}=c_{46}=c_{56}=0$ and $c_{16}=c_{26}=0$. The simpler analysis based on (3) and (4) is therefore valid for any direction of motion of the dislocation line in this plane, and the edge and screw components can be considered separately.

From (49) and (50) the width of the edge component as a function of ρ, s_{ij}' and c is

$$\xi=\tfrac{1}{2}D(c)s_{66}'a_0\left[(1-s_{11}'\rho c^2)-\left\{\frac{s_{11}'(s_{22}'-L\rho c^2)}{(1-s_{66}'\rho c^2)}\right\}^{1/2}\rho c^2\right] \qquad \ldots\ldots(54)$$

where

$$D(c)=[s_{11}'\{s_{66}'+2s_{12}'-\rho c^2(L+s_{11}'s_{66}')+2[s_{11}'(s_{22}'-L\rho c^2)(1-s_{66}'\rho c^2)]^{1/2}\}]^{-1/2}$$

and $L=s_{11}'s_{22}'-s_{12}'^2$. The expression (54) is valid both when the roots of the quartic (9) are real and when they are complex. When $c=0$, (54) becomes

$$\xi=\tfrac{1}{2}Ks_{66}'a_0$$

with

$$K=[s_{11}'\{2s_{12}'+s_{66}'+2(s_{11}'s_{22}')^{1/2}]^{-1/2} \qquad \ldots\ldots(55)$$

which agrees with Eshelby's result (1949 b).

In anisotropic anti-plane strain the vanishing of e_1, e_2, e_3 and e_6 leads to the relation

$$\left.\begin{aligned} p_i&=c_{ij}e_j\\ e_i&=s_{ij}''p_j\end{aligned}\right\} \qquad \ldots\ldots(56)$$

where the suffixes i and j take only the values 4 and 5, and si_j'' is the matrix reciprocal to c_{ij}. In terms of the s_{ij} we have

$$s_{ij}''=\begin{vmatrix} s_{ij} & s_{i1} & s_{i2} & s_{i3} & s_{i6}\\ s_{1j} & s_{11} & s_{12} & s_{13} & s_{16}\\ s_{2j} & s_{21} & s_{22} & s_{23} & s_{26}\\ s_{3j} & s_{31} & s_{32} & s_{33} & s_{36}\\ s_{6j} & s_{61} & s_{62} & s_{63} & s_{66}\end{vmatrix} \div \begin{vmatrix} s_{11} & s_{12} & s_{13} & s_{16}\\ s_{21} & s_{22} & s_{23} & s_{26}\\ s_{31} & s_{32} & s_{33} & s_{36}\\ s_{61} & s_{62} & s_{63} & s_{66}\end{vmatrix} \qquad \ldots\ldots(57)$$

When $c_{14}=c_{24}=c_{15}=c_{25}=c_{34}=c_{35}=c_{46}=c_{56}=0$, p_4 and p_5 are the only non-vanishing stresses. The s_{ij}'' are then equal to the s_{ij}. This is true for a dislocation moving in the (0001) plane in a hexagonal close-packed structure and the expression η in (53) for the width of the screw component may thus be written

$$\eta=(a_0/2s_{55})[s_{44}s_{55}-s_{45}^2-|s_{ij}|s_{55}\rho c^2]^{1/2}=\tfrac{1}{2}a_0[(s_{44}/s_{55})-s_{44}\rho c^2]^{1/2} \qquad \ldots\ldots(58)$$

since $s_{45}=0$. When $c=0$,

$$\eta=\tfrac{1}{2}a_0(s_{44}/s_{55})^{1/2}. \qquad \ldots\ldots(59)$$

From formulae (54) and (58) we can easily derive expressions for the widths in the isotropic approximation. These agree with those published by Eshelby (1949 a).

The widths of the edge and screw components have been computed as a function of dislocation velocity, using the following values for the elastic moduli

s_{ij}^0 (Hearmon 1946): the moduli s_{ij}^0 are referred to standard axes, with x_3^0 along [0001], and are in units of 10^{-13} cm^2 dyn^{-1}.

	Mg	Zn	Cd
s_{11}^0	22·3	7·70	12·3
s_{33}^0	19·8	27·7	35·5
s_{44}^0	59·5	24·4	54·0
s_{12}^0	−7·7	0·45	−1·5
s_{13}^0	−4·5	−6·4	−9·3

Figures 1 and 2 show the results for the edge and screw components respectively. The dislocation widths are in units of a_0, the distance between slip planes in the x_2 direction, and the velocity is in units of $c_2=(\mu/\rho)^{1/2}$ *for each metal*; c_2 is the velocity of shear waves in the isotropic approximation, μ being the isotropic

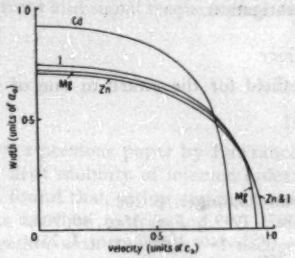


Figure 1.

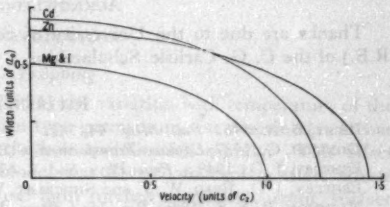


Figure 2.

Widths of uniformly moving dislocations in anisotropic media.

shear modulus and ρ the density. It is convenient first to use $(1/\rho s_{44})^{1/2}$ as the unit of velocity in computing the widths, and then to alter the velocity scale for each metal by multiplying by the factor $(1/\mu s_{44})^{1/2}$. The values used for μ are, in units of 10^{11} dyn cm^{-2}, Mg 1·67, Zn 3·32, Cd 2·40. The curves I are obtained by using the isotropic formulae for the width. All the curves show a monotonic decrease of the width with increasing velocity in this range. Thus there is apparently a numerical error in figure 2 of the published isotropic analysis, which shows a maximum in the width–velocity curve for the edge component (Eshelby 1949 a). In our figures 1 and 2 the curve for isotropy is the same for all three metals, and the difference between the I curve and that for any metal indicates the error made in using the isotropic formulae rather than the anisotropic. As might have been anticipated, this is small for Mg (the Mg and I curves for the screw components are not distinguishable on this scale), but larger for Zn and Cd, where, for example, the limiting velocities for the screw components are higher by factors of 1·44 and 1·23 than those estimated using isotropy. It is, of course, not possible to say how allowance for anisotropy will affect the limiting dislocation speeds attained under given stresses until its effect on the processes causing energy dissipation has been estimated.

The curves for magnesium, zinc and cadmium cannot be compared directly since the velocity scale is different for each metal. To make such a comparison

the velocities must each be multiplied by the factor $(\mu/\rho)^{1/2}$. In units of 10^5 cm sec^{-1} this factor has the values 3·10, 2·19 and 1·67 for magnesium, zinc and cadmium respectively. The resulting absolute limiting velocities are as follows, in units of 10^5 cm sec^{-1}:

	Mg	Zn	Cd
Edge component	2·87	2·08	1·35
Screw component	3·09	3·16	2·05

Direct comparison can, of course, be made of the dislocation widths when $\epsilon = 0$. It is interesting that the widths of both components are greater in cadmium than in magnesium or zinc. Thus, using the simple Peierls–Nabarro formula based on a sinusoidal law for the interatomic forces (Eshelby 1949 b), the shear stress to move a dislocation is less in cadmium than in the other two metals. Whether this distinction would survive the introduction of more realistic laws of force and permit a correlation to be made with the observed yield stresses and cleavage properties is a matter for further investigation.

Acknowledgment

Thanks are due to the University of Sheffield for the award to one of us (R.B.) of the C. G. Carlisle Scholarship.

References

Bilby, B. A., 1953, *Phil. Mag.*, **44**, 782.
Doetsch, G., 1937, *Laplace-Transformation* (Berlin : Springer), p. 202.
Eshelby, J. D., 1949 a, *Proc. Phys. Soc.* A, **62**, 307; 1949 b, *Phil. Mag.*, **40**, 903.
Eshelby, J. D., Read, W. T. and Shockley, W., 1953, *Acta Metallurgica*, **1**, 251.
Hearmon, R. F. S., 1946, *Rev. Mod. Phys.*, **18**, 409.
Nabarro, F. R. N., 1947, *Proc. Phys. Soc.*, **59**, 256.

On the Current Gain of Germanium Filamentary Transistors

By R. LAWRANCE, A. F. GIBSON and J. W. GRANVILLE
Radar Research Establishment, Great Malvern, Worcs.

MS. received 22nd March 1954

Abstract. It has been established by mobility measurements that injected carriers in transit down a germanium rod may be trapped and subsequently released. The trapping of minority carriers is expected to affect materially the performance of germanium devices under certain conditions. The influence of traps on the current gain of germanium filamentary transistors is described in this paper. Measurements of the variation of current gain with temperature, emitter current and signal frequency are described.

§ 1. Introduction

1.1. *Trapping*

In a previous paper by Lawrance (1954) the variation with temperature of the drift mobility of injected holes in n-type germanium was described. It was found that, at low temperatures, the hole mobility fell rapidly. This effect was ascribed to trapping. A trap model was devised to explain these effects together with the effects of background light intensity, emitter current, surface treatment of the specimen, etc. For the sake of completeness this theoretical model will now be described briefly, but familiarity with the original paper is desirable.

The generally accepted theory of contact rectification in semiconductors assumes the existence of additional energy levels on the crystal surface, referred to as surface states, following Bardeen (1947). In an n-type semiconductor these states may capture electrons from the interior of the crystal and a net negative space charge is built up on the surface. This charge raises the potential energy of electrons near the surface and lowers the potential energy for holes. To explain the observed trapping phenomena it is assumed that the surface potential minimum for holes in n-type germanium constitutes a surface trap running the entire length of a germanium rod or filament. This type of trap differs from more conventional models (e.g. F-centres) in some respects, as follows:

(i) The effective trap depth, which can be identified with the height of the potential barrier for electrons at the surface, decreases with increasing hole occupation density. This may be considered to arise from cancellation of the negative surface charge by the positively charged holes, but is discussed more fully by Hogarth (1954).

(ii) The trapped holes may have a small but non-zero mobility along the surface. In practice the hole must travel in a restricted channel between the crystal surface on the one hand and the potential wall on the other. As the width of this channel is small compared with a mean free path the mobility will be very low. If the channel mobility is small compared with the mobility in the body of the material it can for all practical purposes be taken as zero.

(iii) While no direct measurements of the recombination rate of electrons and trapped holes have been made, it is clearly very slow. Experiments by C. A. Hogarth in this laboratory and elsewhere (Bray, Purdue University, private communication) on the decay of photoconductivity in germanium at low temperatures show that holes may be stored for very long periods (of the order of seconds). Additional evidence on this point will be presented later. A trapped hole can therefore be considered as perfectly 'safe' and can only leave the trap by acquiring sufficient thermal energy to re-enter the body of the crystal. This arises, in effect from the sparsity of free electrons in the surface channel. In this respect channels and F-centres are similar.

1.2. *The Current Gain in Transistor Assemblies*

At sufficiently low emitter currents the current gain α in point type transistors rises rapidly to reach very high values. This effect has been described by Sittner (1952) and Cooke-Yarborough and Stephen (1954) for commercially available units. Similar effects have been observed in this laboratory for both formed and unformed collector contacts, in both germanium and silicon. Present theories of the current gain effect (Hogarth 1954) indicate that the value of α is given by

$$\alpha \leqslant 1+\frac{\mu_e}{\mu_p}=1+b=3{\cdot}1 \text{ in germanium of transistor purity}$$

where μ_e and μ_p are the mobilities of electrons and holes respectively. In practice, values of α considerably in excess of 3·1 can be obtained, and Sittner has suggested that trapping, giving rise to a marked reduction in μ_p, might be responsible for the large values observed. That suitable traps exist has been demonstrated, as described above, but the application of this theory to point type transistors is very difficult and is discussed by Hogarth (1954). In the present paper the degree to which trapping affects the current gain in filamentary type transistors will be described, as these devices represent a relatively simple system, amenable to quantitative description. Inasmuch as the measurement of current gain is easier, experimentally, than the measurement of drift mobility, these experiments may be used to check further the theory of trapping presented above.

1.3. *The Filamentary Transistor*

The circuit arrangement of a filamentary transistor is shown in figure 1. The analysis of the filamentary transistor has been given by Shockley, Pearson and Haynes (1949), the relevant results being as follows:

$$\alpha=\frac{R_b}{R_b+R_c}-\frac{R_c\gamma\beta(1+b)}{R_b+R_c}=\alpha_0+(1-\alpha_0)\gamma\beta(1+b) \qquad \ldots\ldots(1)$$

where
$$\beta=\frac{V_c}{K}\left\{1-\exp\left(-\frac{K}{V_c}\right)\right\} \qquad \ldots\ldots(2)$$

and
$$\frac{\tau_t}{\tau_p}=\frac{l_c^2(R_c+R_b)}{\tau_p\mu_p R_c V_c}=\frac{K}{V_c}. \qquad \ldots\ldots(3)$$

γ is injection efficiency of emitter contact ($\gamma \leqslant 1$), β hole transfer efficiency ($\beta \leqslant 1$), V_c collector voltage, τ_t transit time, emitter to collector, τ_p lifetime of injected holes, μ_p mobility of injected holes, R_c collector resistance (filament resistance, emitter to collector), R_b base resistance (filament resistance, emitter to base), l_c length of filament, emitter to collector, $b=\mu_e/\mu_p$.

When $V_c = 0$ the current gain α reduces to α_0 (equation (1)), namely the current gain of the resistive network which would be obtained if there were no transistor action whatsoever.

1.4. *Experimental Details*

Using the circuit shown in figure 1, the current gain was determined by measurement of the ratio of the a.c. voltages appearing across the 10 ohm resistors in the emitter and collector circuits respectively. The input (emitter) signal current was about 0·5 μA r.m.s. superimposed on a variable d.c. bias. In this way the current gain defined by $\alpha = (\partial i_c/\partial i_e)_{V_c \text{ constant}}$ could be determined for any emitter current greater than or about 5 μA. Lower emitter currents could not be used because of the marked rise in emitter impedance at very low levels. Any increase in emitter resistance, together with stray capacitance S, can lead to serious errors at high frequencies. As the filament was mounted in an evacuated micromanipulator which could be cooled (Lawrance 1954), fairly large stray capacities were inevitable.

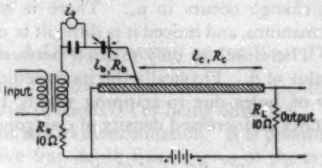


Figure 1.

Except for the measurement of frequency response, a signal frequency of 800 c/s was used throughout. Care was taken to ensure that, even when the mobility was reduced by trapping, a sufficiently high sweeping field was used to keep the transit time short compared with one period of the signal frequency. The transit time was usually comparable with the carrier lifetime in the filament, and β was usually of the order of 0·5.

As a check on the equipment, the variation of α with collector voltage V_c was measured. Using equations (1), (2) and (3) a value of τ_p could be determined from this measurement and compared with the value obtained by more conventional methods. N-type germanium single crystals, with resistivities between 2 and 8 ohm cm, were used throughout.

§ 2. The Effect of Trapping on the Current Gain of Filamentary Transistors

Figure 2 shows the variation of current gain α of a typical filamentary transistor with temperature under various conditions of emitter current and background light intensity. The only quantities in equation (1) that are likely to be affected by trapping are b and β. The probable effect of hole trapping on these quantities will now be discussed.

2.1. *The Effect of Trapping on* β

The value of β is determined by the quantity $l_e^2(R_e + R_b)/\tau_p \mu_p R_e V_c$ where l_e, R_e, R_b and V_c are clearly unaffected by trapping. The average hole mobility is, however, reduced by trapping such that

$$\mu_p = \mu_{p0}\left(\frac{t}{t + t_t}\right), \qquad \ldots\ldots(4)$$

where t_t is the average time spent in a trap and t is the average time between captures in traps. Similarly, if the recombination rate in traps is small, as previously described, the lifetime τ_p is increased by trapping such that

$$\tau_p = \tau_{p0}\left(\frac{t+t_t}{t}\right). \qquad \ldots\ldots (5)$$

Hence the value of the product $\tau_p\mu_p$, and thus β, is unaffected by trapping. It may be relevant to remark at this stage that, in recent experiments carried out in this laboratory with n-type silicon, the increase in lifetime caused by trapping has been measured directly. By an argument similar to that given above it may also be shown that, while trapping can increase the total number of holes (trapped plus untrapped) in the filament, trapping does not affect the distribution of holes.

2.2. *The Effect of Trapping on b*

When trapping of holes occurs, the value of b will increase, due to the fall in μ_p, unless a comparable change occurs in μ_e. There is no evidence of electron trapping in n-type germanium, and indeed it is difficult to envisage any mechanism that might cause it. Therefore the only relation between trapping and current gain is through the value of b. Physically the increased current gain arises from the increased density of holes due to trapping which, to satisfy space charge requirements, necessitates an increased density of electrons of normal mobility.

2.3. *The Effect of Trapping on α*

The above arguments allow some qualitative predictions to be made for comparison with the experimental results given in figure 2, as follows:

(i) Around room temperature, where trapping is not important, it is known that μ_p is proportional to $T^{-2\cdot3}$ and μ_e is proportional to $T^{-1\cdot6}$ (Lawrance 1953, Prince 1953). Hence the value of $1+b$ and therefore the current gain α should decrease monotonically with decreasing temperature. That γ is independent of temperature in this region has been demonstrated by Hogarth (private communication), and β is not expected to vary rapidly with temperature. That α does in fact decrease with decreasing temperature is shown by the results of figure 2, the fall being slightly more rapid than may be accounted for by the change in $1+b$ alone.

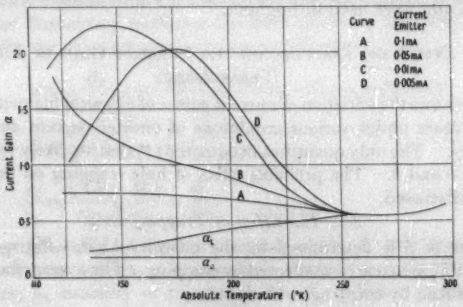


Figure 2.

(ii) At sufficiently low temperatures, when the average time spent in a trap is comparable with the time between captures, the value of μ_p will fall and the current gain will increase markedly. In figure 2 this is observed to occur at about 250°K.

(iii) It was shown by Lawrance (1954) that illumination with white light of sufficient intensity removed the effects of trapping. Similarly the current gain of an illuminated filament decreases with decreasing temperature over the entire range. This is indicated by the curve marked α_L of figure 2 and confirms that the rise in α at low temperatures is due to the trapping of holes.

(iv) By analogy with the effects observed by Lawrance, trapping would be expected to be important at low emitter currents. This feature is illustrated by curves A, B, C and D of figure 2.

(v) Other phenomena observed by Lawrance by drift mobility measurements, for example the effect of the sweeping field magnitude, a transverse magnetic field and the effect of surface treatments, have also been reproduced qualitatively in current gain measurements.

§ 3. A Determination of the Trap Depth

An attempt was made by Lawrance to determine the depth of the traps from drift mobility data, a value of about 0·1 ev being obtained. There were a number of very serious objections to this determination. It is possible, at least in principle, to determine the effective trap depth from curves of the type shown in figure 2. Current gain measurements have one material advantage over mobility measurements as α increases when trapping occurs and measurements become easier. In mobility measurements, however, the collector pulse is markedly rounded off by trapping, making drift times difficult and ultimately impossible to determine.

To determine the depth of the traps we may write

$$\alpha_t=\alpha_0+(1-\alpha_0)\gamma\beta(1+b_t) \qquad \ldots\ldots(6)$$

where the subscript t indicates trapping. On illumination by white light, however, the effects of trapping are removed but β remains unchanged as described above. Hence we may write

$$\alpha_L=\alpha_0+(1-\alpha_0)\gamma\beta(1+b) \qquad \ldots\ldots(7)$$

where the value of b, the mobility ratio without trapping, is known. Thus, dividing equations (6) and (7) to eliminate the temperature dependent variable β and rearranging, we have

$$\frac{b_t}{b}=\frac{\mu_{p0}}{\mu_p}=\frac{b(\alpha_t-\alpha_0)+(\alpha_t-\alpha_L)}{b(\alpha_L-\alpha_0)}=f(\alpha). \qquad \ldots\ldots(8)$$

As α_t, α_L and α_0 are all measured quantities and b is known, the ratio μ_{p0}/μ_p can be determined. It is expected theoretically that

$$\frac{\mu_{p0}}{\mu_p}=1+C\exp(E/\boldsymbol{k}T) \qquad \ldots\ldots(9)$$

where E is the trap depth and C a dimensionless constant variously interpreted by Lawrance (1954) and Sittner (1952). Hence, if the function of α on the right-hand

side of equation (8) is plotted on a logarithmic scale against the reciprocal of the absolute temperature, it is expected that

(*a*) at sufficiently high temperatures such that $1 \gg C\exp(E/\boldsymbol{k}T)$ the value of $f(\alpha)$ will tend to unity: clearly this occurs when $\alpha_t = \alpha_L$, i.e. above 250°K for the unit shown in figure 2;

(*b*) at sufficiently low temperatures such that $1 \ll C\exp(E/\boldsymbol{k}T)$ a straight line of slope $E/\boldsymbol{k}$ should be obtained. The dangers of obtaining activation energies from curves in which temperature is the dependent variable are well known, but no other method of obtaining the trap depth seems to be available unless the constant C can be separately determined.

The experimental results of figure 2 have been replotted in figure 3 for comparison with the above theory, the letters on each curve of figure 3 referring to the

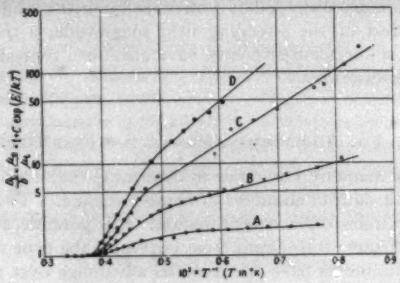


Figure 3.

corresponding curve of figure 2. The expected results (*a*) and (*b*) above are obtained, except that the apparent trap depth decreases with increasing emitter current. The apparent trap depths are as follows:

Emitter current (mA)	0·005	0·01	0·05	0·1
Apparent trap depth (ev)	0·105	0·08	0·033	0·007

If the apparent trap depth can in fact be identified with the potential rise at the surface, as suggested in § 1.1, then a decrease in trap depth with increasing hole density (i.e. emitter current) would be expected. The four points of the table indicate that the trap depth decreases linearly with the logarithm of the emitter current. The value of the slope is 0·075 ev per decade of emitter current. It will be appreciated that as the emitter current is proportional to the added hole density p it follows that

$$\frac{dE}{d(\log i_e)} = \frac{dE}{d(\log p)}$$

and hence the rate of change of the trap depth with hole density is also 0·075 ev per decade.

The variation of the surface barrier height with local hole density is discussed by Hogarth (1954). Hogarth has shown that the maximum theoretical value of $dE_B/d(\log p)$, where E_B is now the barrier height, is $2{\cdot}3\,\boldsymbol{k}T$ or 0·06 ev per decade; experimental values range between 0·02–0·06 ev per decade, depending on hole density. Any quantitative agreement between $dE_B/d(\log p)$ and $dE/d(\log p)$ is certainly partly fortuitous, but the similar magnitude of the two quantities does indicate that the traps and the surface barrier may be identified with each other.

§ 4. The Frequency Dependence of the Current Gain in Filamentary Transistors

In the absence of trapping, and if the sweeping field is not too large, the current gain α should be independent of signal frequency up to some limiting frequency f_0 at which α will fall rapidly (Shockley, Pearson and Haynes 1949). The value of f_0 is given by $f_0 = 1/t_0$, where t_0 is the transit time from emitter to collector. The hole trapping which occurs at low temperatures can be eliminated by illumination and an experimental curve, denoted as α_L, is shown in figure 4.

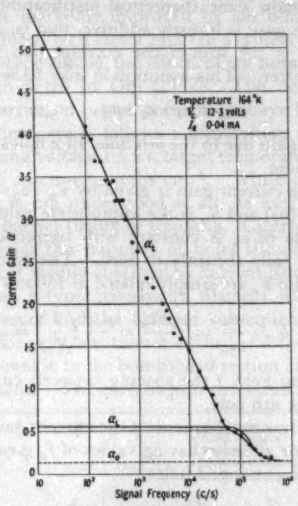


Figure 4.

Clearly the result is as expected. When trapping is important the transit time will be increased, f_0 reduced and the current gain α correspondingly increased. However, the transit time is now no longer single valued and the actual variation of α with frequency will be the sum of many curves of the same form as α_L but with smaller values of f_0 and correspondingly greater values of α. The experimental results shown in figure 4 indicate that α_t, the current gain when trapping is important, is described by the equation

$$\alpha_t = \text{constant}\ (\text{constant} - \log f/f_L) \qquad \ldots\ldots (10)$$

where f_L is some limiting frequency. Inspection of figure 4 shows that this empirical equation is adequate over at least four decades of frequency but does not hold for values of f of the order of f_L. Similarly equation (10) indicates that α_t will be infinite at zero frequency. While this is consistent with the experimental data available, it is clearly an impossible condition to achieve. In practice the value of α_t can only continue to increase at low frequencies if holes with sufficiently long transit times are making a significant contribution to the current

gain. When the transit time is very long (i.e. the time in traps is very long) recombination between trapped holes and free electrons must become important. Holes having transit times in excess of this value will have very low values of β and will make no significant contribution to the current gain. The experimental results indicate that recombination in traps is not important at 5 c/s (the lowest measured frequency) and hence the 'lifetime' of a trapped hole is at least greater than 0·2 second. This is a very large value, considerably in excess of the body lifetime (typically 10–50 μsec), and fully justifies the assumption that trapped holes are completely 'safe' (cf. §§ 1.1 and 2.1).

It is possible to obtain some theoretical justification for equation (10) if it is assumed that current gain is linearly additive, that is, if it is assumed that the observed current gain is the sum of the current gain due to each individual hole injected at the emitter. This assumption may be written as

$$\alpha_{obs} = \alpha_1 + \alpha_2 + \alpha_3 \ldots \alpha_n$$

where α_n is the current gain due to the nth hole, all n holes being assumed injected at the same instant in time.

For the nth hole the current gain is given by $\alpha_n - \alpha_0 = C(1+b_n)$, where C is a constant (see equation (6)) and b_n is the appropriate value of the mobility ratio. Suppose that this value of α_n is constant with frequency up to some limiting frequency f_n where α_n falls abruptly to zero. The frequency f_n and the corresponding mobility ratio b_n are simply related as follows :

$$b_n = \frac{\mu_e}{\mu_{pn}} = \frac{\mu_e F t_n}{l} = \frac{\mu_e F}{l}\left(\frac{1}{f_n}\right) = \frac{A}{f_n}$$

where F is the sweeping field, l the spacing between emitter and collector and t_n the transit time of the nth hole.

Now at some measurement frequency f the current gain α_f will be made up of the sum of all the α_n's for all holes having values of f_n greater than f, up to a limit f_L set by the mobility of untrapped holes.

$$\text{Hence} \quad \alpha_f = \text{constant} \int_{f_L}^{f} \left(1 + \frac{A}{f_n}\right) df_n = \text{constant}\,(f_L - f) - A \log \frac{f}{f_L}. \quad \ldots(11)$$

Equation (11) is not valid for $f=0$ as recombination in traps has been neglected. Similarly it is not valid for f of the order of f_L because of the simplifying assumption that α_n is either constant with frequency or zero. Thus the application of equation (11) should be restricted to frequencies such that $f_L - f \simeq f_L$. If this is done, equation (11) becomes identical with the empirical equation (10). The curve illustrated in figure 4 can be fitted approximately by $f_L = 50$ kc/s, $A = 68$ kc/s and the multiplying constant $7{\cdot}8 \times 10^{-6}$. By definition the value of f_L should be given by

$$b = \frac{\mu_e F}{l}\frac{1}{f_L} = \frac{A}{f_L}. \quad \ldots\ldots(12)$$

From equation (12) and the above values of A and f_L, together with the measured values of F and l, we obtained

$$b = 1{\cdot}35 \quad \text{and} \quad \mu_e = 5{\cdot}7 \times 10^3 \text{ cm}^2 \text{ v}^{-1} \text{ sec}^{-1} \text{ at } 164°\text{K}$$

which are in adequate agreement with the experimental values obtained from drift measurements (Lawrance 1954) at the appropriate temperature.

§ 5. The Effect of Ion Bombardment on the Traps in Germanium

In some originally unrelated experiments, hitherto unpublished, it was found that bombardment of n-type germanium crystals with postive ions reduced the point contact rectification properties to negligible proportions. This indicated that the surface potential barrier had been largely removed. Emitter point efficiency measurements on bombarded surfaces (Hogarth 1953) indicate that the surface barrier height is reduced to unmeasurably small values by the bombardment. Etching the germanium by standard methods completely removes the effects of the bombardment. Conversely, the rectification properties of p-type germanium are markedly improved by ion bombardment, the diode reverse current being reduced by a factor of 20 in a typical case. Similar results are obtained for n- and p-type silicon, the effects of ion bombardment of the latter having been described in detail by Ohl (1952). The optimum conditions for destroying the surface barrier on n-type germanium have been found by experiment and, in our equipment, are as follows : atmosphere nitrogen, gas pressure $\sim 10^{-3}$ mm Hg, accelerating voltage 1·5 kv, target temperature room temperature, total charge delivered 25 micro-coulombs cm^{-2}. While the mechanism of the ion bombardment effect is not understood at all, its existence allows a further check to be made on the barrier theory of trapping. Clearly, if ion bombardment destroys the surface barrier the trapping of holes should also disappear. As indicated in § 3, this is most readily observed by current gain measurements. The current gain of an etched n-type germanium filament was therefore measured as a function of temperature and the filament subsequently bombarded on all four sides. Part of the filament was then re-etched and the emitter put down on this portion, as near as possible to the bombarded section and its original position for the previous experiment. The current gain measurement was repeated.

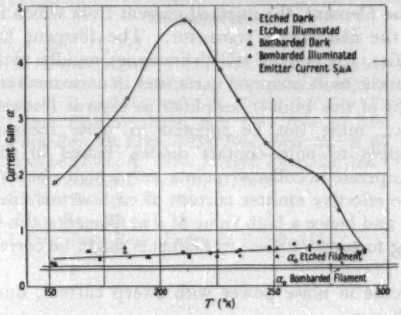


Figure 5.

The results of this experiment are shown in figure 5. It will be seen that the current gain before bombardment follows the usual curve (cf. figure 2, same emitter current). The ' bump ' at about 270°K can probably be dismissed as ice formation as this filament was not measured *in vacuo*. Under illumination, and also after bombardment (both in the dark and illuminated), the current gain falls steadily with decreasing temperature, the experimental results being superimposed within the limits of error. The difference in the values of α_0 for the etched

and bombarded filaments is a measure of the accuracy to which the emitter could be placed on the same point in each case.

The results of this experiment, which has been repeated on other filaments, indicate that the effects of ion bombardment and illumination are electrically the same. Both can be interpreted as reducing the surface barrier height with a consequent elimination of the surface channel, or traps. The effect of illumination taken alone can, however, also be interpreted as saturating traps of the F-centre type with holes, provided that the traps be within a few diffusion lengths of the illuminated surface (Lawrance 1954). As the diffusion length often exceeds the filament thickness, traps in the volume of the material are automatically included. The effects of ion bombardment, however, are confined to a surface layer not more than a few microns thick, and cannot involve the saturation of traps with holes. Hence the combination of these two experiments seems to us to be convincing evidence in favour of hole trapping in the surface barrier.

§ 6. The Importance of Current Gain Data in Noise Measurements

In the study of the frequency and current dependent noise in semiconductors ('current noise') it is becoming increasingly common to study relatively simple systems, such as germanium filaments (Montgomery 1952). In view of this interest it is probably worth while to point out the analogies between noise and current gain measurements. Noise in germanium filaments is measured with essentially the circuit shown in figure 1, except that the emitter is omitted. Injection of holes still takes place, however, due to thermal generation of hole–electron pairs and injection from low lifetime regions of the specimen. Such random injection is believed to contribute a major part of the observed noise (Montgomery 1952). Thus we can imagine that there is a virtual emitter at every point on the filament, the injected current from which is amplified by the current gain of the filamentary transistor. The filament therefore acts as a 'built-in' amplifier, the gain of which increases rapidly with collector voltage (or sweep current, the more common parameter in noise measurements). Clearly the characteristics of this built-in amplifier as regards frequency, applied field, temperature, etc., must not be ignored in noise measurements. Similar considerations apply to point-contact diodes, biased in the high resistance direction, which operate as collector points with a high α at low emitter currents. Furthermore, the effective emitter current of each virtual emitter is very small, so that trapping, and hence a high value of α in filaments, can readily occur.

The following features of noise in filaments might be correlated with current gain phenomena :

(*a*) The increase in noise power with sweep current, due primarily to the increase in β with V_c.

(*b*) The variation of noise power with temperature. This would be expected to be of the same form as the variation in current gain with temperature at very low emitter currents. The curves (for α) at low emitter currents (e.g. figure 2, curve D) are markedly similar to those for noise against temperature shown by Montgomery.

(*c*) If trapping is giving rise to a high α, and hence high noise, illumination should reduce it. Germanium filaments at suitable temperatures have not been examined but the effect has been observed in germanium point-contact diodes by Granville and Gibson (1953).

(*d*) The amplification in a filament will affect the frequency dependence of noise. At room temperature where no significant amount of trapping occurs, the frequency dependence of α will be of the form of α_L in figure 4. The noise output, therefore, would be expected to show an increased rate of fall with frequency when very high frequencies were reached. This may be the explanation of the results of Mattson and van der Ziel (1935) at frequencies of the order of megacycles per second. At low temperatures, when trapping becomes important and the noise is a maximum, the variation of α_t with frequency shown in figure 4 should be significant. No experimental data in this temperature region are available.

§ 7. Conclusions

It is concluded that hole trapping is responsible for the high current gain observed in filamentary transistors at low temperatures and low emitter currents. It would appear that the traps are located on the surface and their effective density can be controlled by ion bombardment and other surface treatments. The available evidence suggests that the depth of the traps decreases with increasing occupation density, a result consistent with the view that the holes are trapped in a surface channel produced by the minimum potential energy for holes near the surface. The mobility of a hole in the channel is believed to be very small and the lifetime of a hole in the channel very long.

Acknowledgments

We are indebted to our colleagues in the transistor group of this establishment for their advice and help. The paper is published by permission of the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office.

References

Bardeen, J., 1947, *Phys. Rev.*, **71**, 717.
Cooke-Yarborough, E. H. and Stephen, J. H., 1954, *Proc. Instn Elect. Engrs*, in the press.
Granville, J. W. and Gibson, A. F., 1953, *Proc. Phys. Soc.* B, **66**, 1118.
Shockley, W., Pearson, G. L. and Haynes, J. R., 1949, *Bell Syst. Tech. J.*, **28**, 364.
Hogarth, C. A., 1953, *Proc. Phys. Soc.* B, **66**, 845; 1954, *Ibid.*, **67**, 636.
Lawrance, R., 1953, *Phys. Rev.*, **89**, 1295; 1954, *Proc. Phys. Soc.* B, **67**, 18.
Mattson, R. H., and van der Ziel, A., 1953, *J. Appl. Phys.*, **24**, 222.
Montgomery, H. C., 1952, *Bell Syst. Tech. J.*, **31**, 950.
Ohl, R. S., 1952, *Bell Syst. Tech. J.*, **31**, 104.
Prince, M. B., 1953, *Phys. Rev.*, **92**, 681.
Sittner, W. R., 1952, *Proc. Instn Radio Engrs, N.Y.*, **40**, 448.

Current Multiplication Processes in n-Type Germanium Point-Contact Transistors

By C. A. HOGARTH

Radar Research Establishment, Gt. Malvern, Worcs.

MS. received 22nd March 1954

Abstract. The various theories that have been put forward to explain the high values of current gain α found for germanium type A transistors, particularly at low values of emitter current, are briefly recounted. Experiments are described which suggest strongly that the 'hole trap' theory as developed by Sittner is not applicable to type A transistors. A possible explanation involving a decrease in collector barrier height by the holes arriving there is examined and is shown to account, at least qualitatively, for the increase in α at low values of emitter current.

§ 1. Introduction

One important parameter of a transistor is the current gain α defined as $(\partial I_c/\partial I_e)_{V_c}$, where I_c and I_e are the collector and emitter currents and V_c is the collector voltage. In general, values of α for type A transistors range from approximately unity to about 2 or 3, but higher values have been reported. The occurrence of high values of α has been the subject of considerable experimental investigation and has excited much interest, particularly in so far as a large increase in α at low values of I_e is observed.

§ 2. Theories of High Values of α

The original explanation of high α values ($\alpha>1$) was given by Bardeen and Brattain (1949) and involved a modulation of space charge at the collector barrier by the holes arriving there. By this means more electrons could flow into the germanium at the collector and hence values of α in excess of unity could be expected.

Shockley (1950) proposed two theories, the 'p–n hook' theory and the 'hole trap' theory. Experiments described by Valdes (1952) suggest that the p–n hook theory does explain the results obtained with type A transistors in which the collector has been heavily electroformed by, for example, pulsing electrically a whisker contact containing a Group V element. Hunter, Fleisher and Irish (1952) have shown that, with formed collectors, effects seemingly characteristic of both of Shockley's theories may occur simultaneously, i.e. large values of α at large values of I_e, together with a rise in α at low values of I_e. Sittner (1952) has applied the theory of a filamentary transistor with traps directly to the type A transistor, and since the controlling term in the current gain of a filamentary transistor is $1+b$, where b is the mobility ratio, any decrease in hole mobility due to trapping would increase b and hence α. For low values of I_e the proportion of holes trapped is very high and α is large; for larger values of I_e the traps are saturated and α is reduced.

The experiments on the temperature variation of drift mobility of holes (Lawrance 1953, 1954) and on current gain in filamentary transistors (Lawrance, Gibson and Granville 1954) are entirely consistent with a trapping model. High

values of α in filamentary transistors occur for low values of I_e at those temperatures at which trapping processes affect the drift mobility of holes. To assess whether or not the theory of filamentary transistors may be applied to type A assemblies we compare known properties of the two systems.

(i) *Effect of background illumination.*

White light reduces the high current gain at low values of I_e for both cooled germanium filamentary transistors and for type A transistors at room temperature or cooled. This may be simply explained by a trapping model but is in fact inevitable, since the holes injected optically bias the emitter to effectively higher values of I_e and hence to lower values of α (Granville and Gibson 1953).

(ii) *Room temperature effects.*

In no experiment on hole drift mobility or the current gain in filamentary transistors have effects of trapping been observed at room temperature. The effects of high α values at low values of I_e may be readily observed in type A transistors at room temperature. In this respect the two assemblies are quite different.

(iii) *Frequency response.*

The frequency response measurements of Lawrance, Gibson and Granville (1954) on filamentary transistors indicate that, when trapping is important, the high value of α for low values of I_e is frequency dependent. A number of type A transistors giving α values between 25 and 50 at low emitter currents were kindly examined at our request by Dr. J. H. Stephen of the Atomic Energy Research Establishment. No significant change in α between 100 and 5000 c/s was observed.

Thus the properties of type A and filamentary transistors differ significantly in at least two major respects. In the following experiments, which attempt to resolve this difficulty, the collector contacts used were of tungsten and unformed, so as to eliminate the possibility of p–n hook formation.

§ 3. Filamentary Studies with the Wide-Spaced Transistor

In order to assess what correlation, if any, exists between the high values of α in filaments and in point-contact transistors, some filamentary assemblies were made from well-etched 7 ohm cm single crystal germanium. The filaments were of cross section about 0·5 mm × 0·5 mm and the emitter and collector contacts were from 3 to 7 mm apart. The assembly could be investigated either as a filamentary transistor or as a wide-spaced type A transistor in which the injected holes travel some distance along the filament before reaching the collector. The test frequency in these experiments was 800 c/s.

Figure 1 shows the effect of sweeping current on α_e, the current gain of the wide-spaced transistor. Initially the sweeping field has the same effect as reducing the distance between the emitter and collector, but for larger sweeping currents α_e decreases since the injected holes tend to be swept out at the end of the filament rather than being collected at the collector contact. α_e is obviously much less (about 400 times) than the α for a normal type A assembly because of (*a*) hole–electron recombination and (*b*) hole spreading. The value of α_e can nevertheless be taken as an indication of the corresponding value of α for normal transistor spacings.

Figure 2 shows the values of α_f for the filament and α_c for the collector measured with the same emitter contact at room temperature. α_c for the collector rises at low values of I_e just as in the type A transistor, but the gain in the filament remains practically constant. Thus the experiment demonstrates that the rise in α_c is a feature only of the collector contact and is not associated with the type of traps discovered by Lawrance (1953, 1954), since these traps are not effective at room temperature. Some other mechanism must therefore be postulated for high values of α at low values of I_e when collector contacts are involved.

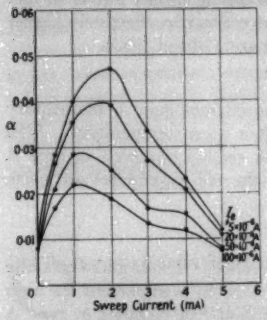


Figure 1. The current gain of a wide-spaced transistor as a function of sweeping current.

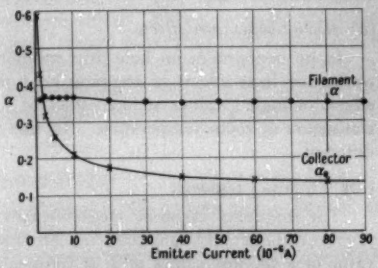


Figure 2. Comparison of current gain values for filamentary and wide-spaced transistors (same filament, sweep current and emitter, room temperature).

§ 4. Collector Current and Local Hole Density

As a means of studying the variation of α_c for a wide-spaced transistor with P_0 the added local hole density near the collector, an assembly similar to those described in § 3 was made and non-rectifying potential probes P_1, P_2, of 0·002 inch platinum wire, were micro-soldered on either side of the collector and about 0·3 mm apart (the technique of making such contacts is described by Mitchell 1954). When forward current I_e was passed through the emitter the collector current could be measured under known operating conditions of sweeping field, collector voltage and load, and also the change in potential at P_1, P_2 for the same total current through the filament. Then P_0 could be determined for the small element near the collector from the equation

$$P_0 = \Delta G A / l e \mu_h (1+b) \qquad \ldots\ldots (1)$$

where A is the area of cross section of the filament, l is the distance P_1P_2, ΔG is the charge of conductance due to P_0, and μ_h is the hole mobility. Thus for given operating conditions I_c and P_0 may be related. In practice P_0/N_0 is usually plotted, N_0 being the normal free electron concentration at the measurement temperature. So as to depart as little as possible from equilibrium conditions in the semiconductor, I_e was adjusted so that P_0 near the collector was at least an order of magnitude smaller than P_n, the intrinsically generated hole concentration.

Figure 3 shows the change in I_c with P_0/N_0 to be linear for fairly large values of P_0/N_0, but for smaller values I_c decreases more rapidly. It should be noted that

the form of figure 3 is not in complete agreement with the results of Shockley, Pearson and Haynes (1949) in a similar type of experiment. They used the actual conductance as measured between two collector points to determine the hole

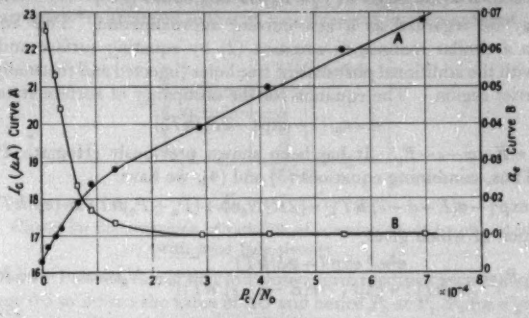


Figure 3. Experimental curves relating local hole density with A, collector current; B, α_e.

concentration, and such contacts are known to give rise to floating potentials (Bardeen 1950) which may affect the measured potential difference. The non-rectifying probes used in our experiments show no injection and hence no floating potentials, so that this measuring technique is considered more reliable for the determination of P_0/N_0, particularly the lower values. In addition there is considerably less noise in the non-rectifying contacts. From the slopes of figure 3 values of α_c were calculated for values of P_0/N_0 corresponding to given values of I_e, and these are also plotted in figure 3. The variation α_c with P_0/N_0 is of the same form as the variation of α with I_e for type A transistors.

In view of our conclusion in § 3 that trapping of the kind that is effective in filamentary transistors is not responsible for the high α values at low values of I_e when collector contacts are involved, an alternative explanation is required. We shall postulate that the increased α_e results from an increased electron current due to some modulation of the barrier by the holes arriving there. Several models could be devised to give an increased electron current. We shall consider a *decrease* in barrier height by the holes arriving at the vicinity of the collector.

§ 5. Variation in Surface Barrier Height with Local Hole Density

5.1. *Theoretical*

We treat in a simple manner the surface state Schottky barrier on n–type germanium. n_0, n is the density of available (occupied) surface states, ϕ (v) is the surface potential rise, E(v) the forbidden energy gap, P_n the concentration of thermally excited holes, P_0 the concentration of injected holes in the collector region, K the dielectric constant of germanium, ϵ the Fermi energy, N_0 the concentration of impurities (assumed fully ionized at room temperature), E_d the depth of surface states below the Fermi level, E_s the height of surface states above the filled band.

Bardeen (1947) has shown that the surface state density and surface potential rise are related by an equation of the form

$$n=(2KN_0\phi/e)^{1/2}. \qquad \ldots\ldots (2)$$

During injection ($P_0 \neq 0$) it may be shown (Mrs. A. M. Woodward, private communication) that, so long as P_0 is small, equation (2) should be replaced, with small remaining error, by the expression

$$n=[2K\{N_0e\phi+(P_n+P_0)kT\exp(e\phi/kT)\}/e^2]^{1/2} \quad \ldots\ldots(3)$$

which may be regarded as a zero-current approximation. This equation is derived in a similar manner to equation (2) by equating surface and internal charges, with the additional postulate of free holes (injected and thermally excited) in the barrier region. The equation for the occupancy of surface states is

$$n=n_0/\{1+\exp(-eE_d/kT)\} \quad \ldots\ldots(4)$$

where $E_d=E-\phi-\epsilon-E_s$. It has been shown previously (Hogarth 1953) that $E_s=0$. Thus, combining equations (3) and (4), we have

$$n_0/[1+\exp\{-e(E-\phi-\epsilon)/kT\}]=[2K\{N_0e\phi+(P_n+P_0)kT\exp(e\phi/kT\}/e^2]^{1/2},$$

simplification of which gives

$$P_n+P_0=\frac{e^2n_0^2\exp(-e\phi/kT)}{2KkT[1+\exp\{-e(E-\phi-\epsilon)/kT\}]^2}-N_0e\phi\exp(-e\phi/kT). \quad \ldots\ldots(5)$$

For germanium at 290°K, to a good degree of approximation,

$$[1+\exp\{-e(E-\phi-\epsilon)/kT\}]^2=1.$$

Thus equation (5) becomes

$$P_0=(1/kT)\exp(-e\phi/kT)[e^2n_0^2/2K-N_0e\phi]-P_n. \quad \ldots\ldots(6)$$

We now investigate the derivative $d\phi/d\log P_0$.

From equation (6) we have

$$\frac{d\log P_0}{d\phi}=-\frac{1}{kT}\left[\frac{N_0e\exp(-e\phi/kT)+e/kT\exp(-e\phi/kT)[e^2n_0^2/2K-N_0e\phi]}{(1/kT)\exp(-e\phi/kT)[e^2n_0^2/2K-N_0e\phi]-P_n}\right]$$

which simplifies to

$$\frac{d\phi}{d\log P_0}=-\frac{kT}{e}\left[\frac{1-\{kTP_n\exp(e\phi/kT)\}/\{e^2n_0^2/2K-N_0e\phi\}}{1+kTN_0/\{e^2n_0^2/2K-N_0e\phi\}}\right]. \quad \ldots\ldots(7)$$

It may be seen that the left-hand side of equation (7) is equal to $d\phi/d(\log P_0/N_0)$ since N_0 is constant.

By a careful inspection of equation (7) and the original equations (2) and (3) it may be shown that the maximum value of $d\phi/d(\log P_0/N_0)$ is $-kT/e$ and the minimum value is zero.

5.2. *Experimental*

The experimental test of equation (7) uses the fact that the injection efficiency γ of a point contact can be related to the surface potential rise ϕ (Hogarth 1953). Thus a decrease in γ implies a decrease in ϕ and vice versa. The general circuit diagram is shown in figure 4. The arrangement is, in effect, a method of measuring the efficiency of emitter E_2 when E_1 is used to inject holes into the filament so that there will be a measurable hole density in the vicinity of E_2 due to E_1. The etched filament was mounted as described previously and four non-rectifying probes P_1 P_2 P_3 P_4 arranged in two pairs were micro-soldered to it (Mitchell 1954). Emitter contacts of 0·002 inch tungsten wire were made to the filament by means of a micromanipulator.

The distances and dimensions of the specimen were measured with a travelling microscope. The measuring procedure was as follows. (i) With no current

through E_1 measure the value of γ for E_2 using a small emitter current I_{E2} of order $20\text{–}30 \times 10^{-6}$ A. (ii) With constant currents of $0{\cdot}1, 0{\cdot}2 \ldots \times 10^{-3}$ A through E_1 measure γ for E_2 and also the change of conductance ΔG at P_1, P_2. Then from equation (1) the value of P_0 in the vicinity of E_2 could be estimated. The emitter currents were arranged so that P_0 was always at least an order of magnitude

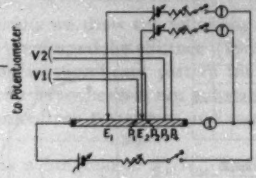


Figure 4. Circuit for determination of variation of injection efficiency of a whisker contact with local hole density.

smaller than P_n. It was found that to a close approximation for sweeping currents in the range $0{\cdot}5$ to $2{\cdot}5$ mA the value of ΔG and hence P_0 at P_1, P_2 for a given small I_{E1} was constant. Since the sweeping current has to be varied in a measurement of γ so as to vary the transit time, this fact provided a useful experimental simplification. The concentration of free electons N_0 was calculated from the measured unperturbed resistivity of the single crystal germanium. The resistivity was $7{\cdot}4$ ohm cm, N_0 was $2{\cdot}35 \times 10^{14}$ cm^{-3}, and ϵ was $0{\cdot}29$ ev.

5.3. *Results*

Experiments were carried out on two filaments of 7 ohm cm germanium. Results for the measurement of γ with I_{E1} are given in figure 5 and it is seen that γ falls sharply for low values of I_{E1} and above a fairly small value of I_{E1} falls more slowly. An approximate theory has been given for the calculation of ϕ from measurements of γ (Hogarth 1953). A more rigorous treatment using quasi-Fermi levels has been derived by Gunn (1954) for the calculation of hole and electron currents separately at a semiconductor–metal contact. Gunn's treatment leads to an equation for the ratio of hole to electron current which is of similar form to that derived from the previous theory but differing by a numerical factor. Values of ϕ calculated from measurements of γ and from either theory lead to similar results, although in general the values obtained from Gunn's theory are rather higher. If we postulate that the reduction in the value of γ observed experimentally is a direct consequence of the increased electron current, then we may correlate changes in γ with changes in ϕ. This has been done by means of the previous theory, and the results for ϕ as a function of P_0/N_0 are given in figure 6. The slopes of the curves vary with P_0/N_0 but the maximum slope $d\phi/d\,(\log (P_0/N_0))$ is $-2{\cdot}2 \times 10^{-2}$ volt, just less than the room temperature value of $-\boldsymbol{kT}/\boldsymbol{e}$, predicted above as the maximum value of the variation. Thus agreement between theory and experiment is obtained, and from the calculated values of ϕ and I_e we may calculate the value of I_c for given values of ϕ and hence construct theoretical curves relating α_e with I_e or P_0/N_0.

§ 6. Current Gain in Terms of Barrier Height Modulation

The reverse current through E_2 now regarded as a collector, calculated from the derived value of ϕ for a given P_0/N_0, is shown in figure 7 as a function of

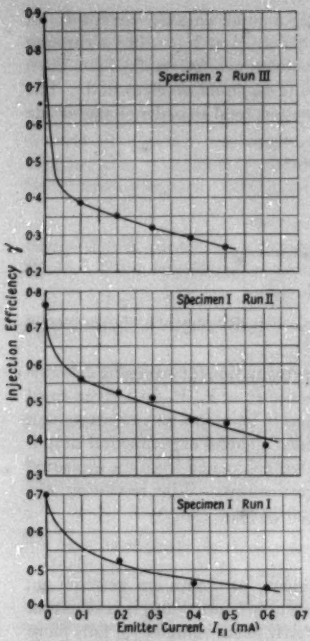


Figure 5. The variation of injection efficiency of E_2 with forward current through emitter E_1.

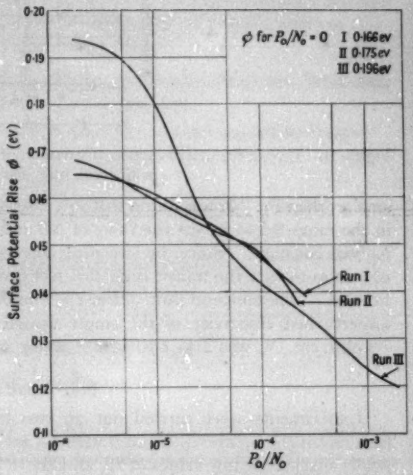


Figure 6. Derived variation of surface potential rise with local hole density for two germanium specimens.

P_0/N_0. The results from run III of figure 6 are used as these cover the largest range of P_0/N_0. The reverse current is calculated from the expression $I_{c0} \exp[-e(\phi+\epsilon)/kT]$, which may be written as $A \exp(-e\phi/kT)$. In view of the uncertainty concerning the particular value of I_{c0} to use in the calculation the

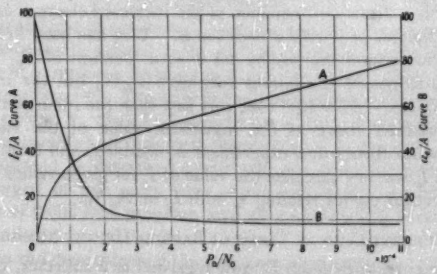


Figure 7. Theoretical curves relating local hole density with A, collector current; B, α_e.

reverse current is plotted arbitrarily as I_e/A. The variation of I_e/A with P_0/N_0 is seen to be of the same form as that for I_e with P_0/N_0 for the wide-spaced transistor (figure 3). The corresponding variation of α_e/A with P_0/N_0 (figure 7) is also similar in form to the experimental curve (figure 3).

§ 7. Conclusions

From the preceding results we draw the following conclusions:

(i) Traps of the kind envisaged by Sittner (1952), Lawrance (1954) and Lawrance *et al.* (1954) play no significant part in the origin of current gain in type A transistors. Sittner's hypothesis is not substantiated because of the lack of correlation between filamentary transistors and assemblies carrying collector contacts.

(ii) The curves of α_e as a function of P_0/N_0, derived from values of ϕ as a function of P_0/N_0, indicate that a theory involving barrier height modulation provides a sufficient explanation for high values of α at low values of I_e.

(iii) From space charge considerations (Shockley 1950) the value of α arising from the above mechanism cannot exceed $1+b$, where b is the ratio of two microscopic mobilities, μ_e/μ_h. In order to account for the values of α greater than $1+b$ it must be assumed that μ_h is reduced immediately at the collector contact. This does not affect statement (ii) since a reduction in μ_h would simply mean that ϕ was reduced by more than the limit set by the normal value of $1+b$.

(iv) Any reduction in μ_h at the contact is essentially a microscopic phenomenon taking place at the semiconductor–metal interface and not in any way characteristic of the semiconductor. Any attempt to explain the decrease in μ_h must inevitably be arbitrary, and it is difficult to conceive any method of verifying it experimentally. The difficulty is emphasized still more by the constant value for drift mobility observed on the same filament with different collector contacts.

Acknowledgments

The author is indebted to his colleagues for much help and advice, and in particular to Dr. A. F. Gibson for many discussions regarding the theory, to Messrs. R. Lawrance and N. Wilson (of Edinburgh University) who collaborated in the measurements reported in § 3, and to Mr. W. H. Mitchell for assistance in mounting some specimens. He is also indebted to the Chief Scientist, Ministry of Supply, and to the Controller, H.M. Stationery Office, for permission to publish this paper.

References

Bardeen, J., 1947, *Phys. Rev.*, **71**, 717; 1950, *Bell Syst. Tech. J.*, **29**, 469.
Bardeen, J., and Brattain, W. H., 1949, *Phys. Rev.*, **75**, 1208.
Granville, J. W., and Gibson, A. F., 1953, *Proc. Phys. Soc.* B, **66**, 1118.
Gunn, J. B., 1954, *Proc. Phys. Soc.* B, **67**, 575.
Hogarth, C. A., 1953, *Proc. Phys. Soc.* B, **66**, 845.
Hunter, L. P., Fleisher, H., and Irish, C. G., 1952, *Trans. Inst. Radio Engrs, N.Y.*, Prof. Group on Electron Devices, No. 1, 25.
Lawrance, R., 1953, *Phys. Rev.*, **89**, 1295; 1954, *Proc. Phys. Soc.* B, **67**, 18.
Lawrance, R., Gibson, A. F., and Granville, J. W., 1954, *Proc. Phys. Soc.* B, **67**, 625.
Mitchell, W. H., 1954, *J. Sci. Instrum.*, **31**, 147.
Shockley, W., 1950, *Phys. Rev.*, **78**, 294.
Shockley, W., Pearson, G. L., and Haynes, J. R., 1949, *Bell Syst. Tech. J.*, **28**, 344.
Sittner, W. R., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 448.
Valdes, B., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 445.

RESEARCH NOTES

The Resistance–Temperature Characteristics of Leaded Brass Wires at Helium Temperatures

By D. H. PARKINSON and J. E. QUARRINGTON
Radar Research Establishment, Malvern, Worcs.

MS. received 29th April 1954

§ 1. Introduction

There have been a number of investigations into the low temperature properties of alloys containing small quantities of lead and other superconducting elements, e.g. by Keesom and van den Ende (1929), Babbitt and Mendelssohn (1935). Phosphor bronze wires have been used very extensively as thermometers in the helium temperature range. More recently leaded brass wires as suggested by Mendelssohn have been used successfully for calorimetric purposes. The temperature at which these wires become sensitive, and also the temperature sensitive range, depends on the quantity and state of the added superconducting element, the degree of work-hardening or strain in the material and on the diameter of the wire. In general thinner wires show sensitivity at lower temperatures.

Two years ago a good supply of leaded brass wire 0·004 in. in diameter and of composition 62% Cu, 36% Zn, 1·73% Pb and 0·08% Ni was obtained from Dr. Chaston of Messrs. Johnson Matthey's research laboratory. This showed sensitivity between approximately 4°K and 5°K only. Wire of 0·002 in. diameter and of the same composition showed sensitivity from 2°K downwards. The shape of the resistance curve and narrowness of the sensitive range was unsatisfactory for normal thermometric purposes. Rather than attempt to alter the characteristics by redrawing, the effect of annealing in a manner similar to that used by Van Dijk (1951) was examined. The results were fairly satisfactory, and wires sensitive over a wider range have been produced. Numerous data have been accumulated on their behaviour after annealing; these are summarized below.

§ 2. Experimental

Annealing was carried out with suitable lengths of the wires wound round formers of pyropholite placed in a temperature-controlled oven. Both the resistance of the specimens and the annealing temperatures were continuously recorded using a two-channel Cambridge recording voltmeter. The control was within ±1°C using temperatures between 160°C and 400°C. The normal annealing time was 24 hours and the specimens were cooled rapidly to room temperature. They were then wound on the cylindrical surface of thin copper drums, insulated and held in place with Araldite varnish. The resistance–temperature characteristics were then investigated by mounting the drum in a liquefier cryostat combination (Parkinson 1954) and condensing a quantity of liquid helium inside it. This liquid was then used for vapour pressure calibrations. For that characteristic which lies above 4·2°K the drum was made

the bulb of a constant volume gas thermometer using a filling of 15 cm of helium gas at 4·2°K. Electrical measurements were made using a Wheatstone bridge with fully compensated leads and 1 : 1 ratio arms.

§3. Results

The useful results are summarized in figures 1 and 2. For all specimens the reduced resistance R/R_{20} is shown (R_{20} is the value of R at 20°K). With wires of 0·004 in. diameter and for annealing temperatures between 195°K and 240°K the characteristic is moved very nearly parallel to itself and to lower temperatures. From approximately 240°K to 310°K it moves back somewhat as shown in figure 2, and from 325°K upwards the characteristic continues to be moved to lower temperatures and widened. A brief investigation showed that wires of 0·002 in. diameter have a similar behaviour (see figure 1). These studies have been carried out largely as thermometers have been required, and are consequently still somewhat incomplete.

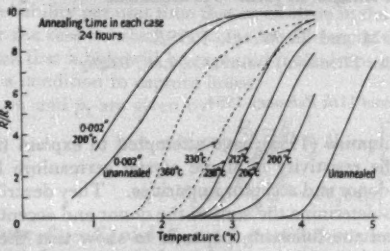


Figure 1. Reduced resistance–temperature characteristics of leaded brass wires annealed between 195°c and 240°c and above 325°c.

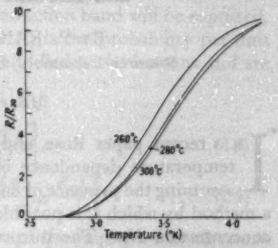


Figure 2. Reduced resistance–temperature characteristics of leaded brass wires annealed between 240°c and 310°c.

An x-ray crystallographic examination carried out by Miss A. Rennie showed that the original specimen contained only an α brass phase. Those annealed between 200°c and 320°c showed the additional β phase, which is not apparent in wires annealed at over about 350°c. In all instances lines corresponding to lead were seen (face-centred cubic $\alpha_0 = 4{\cdot}93 \pm 0{\cdot}01$) with some evidence of preferred orientation. There was also evidence of considerable strain which was relieved by the annealing process, particularly at higher temperatures. It is evident that the behaviour of the wires is related to the phase diagram for brass. The composition used is almost on the boundary separating the α phase from the mixed $\alpha+\beta'$ phases at 300°c. It is probable that a brass in which less zinc is used would show characteristics which are all altered in the same direction by increasing the annealing temperature.

By choosing two wires from the family shown and using them in series it is possible to have a thermometer of good sensitivity over the whole temperature range from 1°K to 4·2°K: the average value of $1/R \times dR/dT$ at the centre of the characteristic of a single wire is approximately 1·4.

These wires seem to have quite stable characteristics but, of course, suffer from the usual disadvantages of dependence on current and magnetic field. The results shown were all found with a measuring current of 0·1 mA.

ACKNOWLEDGMENTS

The authors wish to thank Dr. J. C. Chaston for providing the original supplies of wires and Miss A. Rennie for her ready cooperation in carrying out the crystallographic studies. This note is published by permission of the Chief Scientist, Ministry of Supply, and of the Controller, H.M. Stationery Office.

REFERENCES

BABBITT, J. D., and MENDELSSOHN, K., 1935, *Phil. Mag.*, [7], **20**, 1025.
KEESOM, W. H., and VAN DEN ENDE, J. N., 1929, *Leiden Comm.*, Nos. 203 c and 209 b.
PARKINSON, D. H., 1954, *J. Sci. Instrum.*, **31**, 178.
VAN DIJK, H., 1951, *Proceedings of the Oxford Conference on Low Temperature Physics*, p. 49.

On the Relation between the Sum of Donor and Acceptor Concentration and Lifetime in Single Crystal Germanium

BY P. RANSOM and F. W. G. ROSE
Research Laboratory, British Thomson-Houston Co. Ltd., Rugby

MS. received 11*th February* 1954

IN a recent paper Rose and Timmins (1953) have attempted to explain the temperature dependence of the resistivity of single crystal germanium by assuming the presence of both donor and acceptor impurities. They describe a method by which it is possible to determine the sum of the donor and acceptor concentration N_I. The purpose of the following paper is to show that there is a relation between N_I, the activation energy of the traps and the lifetime of minority carriers, in n-type single crystal germanium.

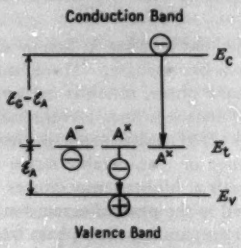


Figure 1. Trapping of electrons and holes by acceptor traps.
$A^\times$ are neutral traps.
A^- are negatively charged traps.

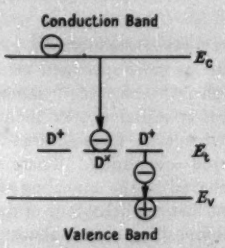


Figure 2. Trapping of electrons and holes by donor traps.
$D^\times$ are neutral donors.
D^+ are donors which have lost their electron and are thus positively charged.

All traps for electrons and holes may be considered as types of donors or acceptors, with various values of activation energies. Two types of trapping mechanism are illustrated in figures 1 and 2, where E_c is the lowest energy level for electrons in the conduction band, E_v is the highest energy level in the valence

band, and E_t is the energy level of the traps. From these figures it is evident that traps can be subdivided into 'donor traps' and 'acceptor traps', and can thus be included in the values of the total impurity concentration N_I. Since the lifetime of minority carriers is a function of the trap concentration, some relation between the lifetime and N_I is to be expected.

A basis for this relation is given by a theory of the "Statistics of the Recombination of Holes and Electrons" (Shockley and Read 1952). The lifetime of holes from equation (A 7) of this reference is given by

$$\tau_p = \frac{\tau_{n0}(p_0+p_1)+\tau_{p0}[n_0+n_1+N_t(1+n_0/n_1)^{-1}]}{n_0+p_0+N_t(1+n_0/n_1)^{-1}(1+n_1/n_0)^{-1}} \qquad \ldots\ldots(1)$$

where N_t is the concentration of traps, n_0, p_0 the equilibrium concentrations of free electrons (free holes), n_1, p_1 the constants of the mass action law for the process of trapping electrons (holes) = concentration of electrons (holes) in the conduction (valence) band when the Fermi level coincides with E_t, n, p the concentration of electrons (holes), $\tau_{n0} = n/R_n$, $\tau_{p0} = p/R_p$, where R_n is the probability per unit time that an electron in the conduction band will be captured, for the case in which the traps are all empty, and R_p is the probability per unit time that a hole will be captured if the traps are all filled with electrons and are in a condition to capture holes.

n_1 and p_1 are given by

$$p_1 = N_v \exp\frac{E_v - E_t}{kT} = N_v \exp\left[\frac{-\mathscr{E}_A}{kT}\right], \qquad \ldots\ldots(2a)$$

$$n_1 = N_c \exp\frac{E_t - E_c}{kT} = N_c \exp\left[\frac{-(\mathscr{E}_G - \mathscr{E}_A)}{kT}\right] \qquad \ldots\ldots(2b)$$

where $\mathscr{E}_G - \mathscr{E}_A$ and $\mathscr{E}_A$ are the activation energies of the traps for free electrons and holes respectively and where N_c is the effective density of energy levels for electrons in the conduction band and N_v is the effective density of energy levels in the valence band: $N_c \simeq N_v \simeq 4{\cdot}82 \times 10^{15} T^{3/2}$.

The probability per unit time that a free electron becomes trapped is proportional to the concentration of free traps and to the concentration of free electrons. A similar argument applies to the case of a hole in the valence band; thus, if acceptor traps are assumed,

$$\tau_{n0} = \frac{n}{R_n} = \frac{n}{\alpha_n N_{ta} n} = \frac{1}{\alpha_n N_{ta}} \qquad \ldots\ldots(3a)$$

and

$$\tau_{p0} = \frac{p}{R_p} = \frac{p}{\alpha N_{ta} p} = \frac{1}{\alpha N_{ta}} \qquad \ldots\ldots(3b)$$

where α_n is the recombination constant for neutral traps and α is the recombination constant for charged traps. Since the trapping rate by charged centres is likely to be greater than by neutral centres, α_n will most probably be smaller than α. By substituting equation (3) in equation (1)

$$\tau_p = \frac{\frac{1}{\alpha_n N_{ta}}(p_0+p_1)+\frac{1}{\alpha N_{ta}}[n_0+n_1+N_{ta}(1+n_0/n_1)^{-1}]}{n_0+p_0+N_{ta}(1+n_0/n_1)^{-1}(1+n_1/n_0)^{-1}}. \qquad \ldots\ldots(4)$$

The lifetime τ_p in equation (4) can be extremely sensitive to variations in temperature for certain values of n_0, p_0, n_1, p_1 and N_{ta}, since n_1 and p_1 from equation (2) are very temperature dependent. If a plot of $\log\tau_p$ against $1/T$

gives a straight line, it may be assumed that either n_1 or p_1 determines the temperature dependence, and the activation energy $\mathscr{E}_A$ or $\mathscr{E}_G - \mathscr{E}_A$ may be calculated from the slope. Such an evaluation has been proposed by Hall (1952), and he obtains a value $\mathscr{E}_A = 0{\cdot}22$ ev from his measurements on a sample of n-type germanium having a resistivity of 8 ohm cm.

With a value of $\mathscr{E}_A \leqslant 0{\cdot}22$ ev for n-type samples we can assume $n_1 \ll n_0$, $p_0 \ll p_1$, $p_0 \leqslant n_0$, and equation (4) reduces to

$$\tau_p \simeq \frac{p_1}{\alpha_n N_{ta}(n_0 + p_0 + N_{ta} n_1/n_0)} + \frac{n_0 + N_{ta} n_1/n_0}{\alpha N_{ta}(n_0 + p_0 + N_{ta} n_1/n_0)}. \quad \ldots\ldots(5)$$

Since $\alpha_n < \alpha$, if $N_{ta}(n_1/n_0)$ is sufficiently small and $n_0 \ll p_1$, equation (5) can be approximated to the form

$$\tau_p \simeq \frac{p_1}{\alpha_n N_{ta}(n_0 + p_0)} = \frac{N_v \exp(-\mathscr{E}_A/kT)}{\alpha_n N_{ta}(n_0 + p_0)}. \quad \ldots\ldots(6)$$

The concentration N_{ta} of acceptor traps can be calculated from the total impurity concentration N_I if all the acceptors are considered as traps having the same activation energy:

$$N_I = N_D + N_{ta} \quad \ldots\ldots(7)$$

where N_D is the concentration of donors.

Since more than 99% of the traps will be charged, the condition for neutrality in n-type or intrinsic material is

$$n_0 - p_0 \simeq N_D - N_{ta}. \quad \ldots\ldots(8)$$

From equations (7) and (8)

$$N_{ta} = \tfrac{1}{2}(N_I - n_0 + p_0). \quad \ldots\ldots(9)$$

Thus N_{ta} can be calculated from equation (9). The value p_0 which is given by $p_0 = 1{\cdot}96 \times 10^{26}/n_0 \leqslant 1{\cdot}4 \times 10^{13}$ cm^{-3} can be neglected with all extrinsic samples.

With intrinsic samples $p_0 - n_0$ can be neglected. From equations (6) and (9) we have

$$\tau_p \simeq \frac{2N_v \exp(-\mathscr{E}_A/kT)}{\alpha_n(N_I - n_0 + p_0)(n_0 + p_0)}. \quad \ldots\ldots(10)$$

We have made lifetime measurements in the temperature range 242°K to 302°K using the Morton–Haynes method (Valdes 1952) on samples of antimony-doped n-type germanium. We have also made measurements in the range 242°K to 319°K on two undoped samples (VU1 and VU2 in the table) which were intrinsic above 291°K and extrinsic n-type at lower temperatures. The type of conductivity was determined from measurements of Hall coefficient and resistivity at various temperatures. Since the Morton–Haynes method only applies to the measurment of the lifetime of minority carriers, the measurements on the intrinsic samples are not very reliable.

The activation energies of the traps have been calculated from the experimental curves shown in figure 3 and are tabulated as 'experimental $\mathscr{E}_A$ values, in the table. In this table also the total impurity concentrations N_I and the resistivities ρ at 291°K are given. The values of N_I have been determined from resistivity measurements at various temperatures.

It must be borne in mind when considering the resistivity values given in the table that the resistivity is determined mainly by the difference between donor and acceptor concentrations, but is modified also by the sum N_I of these concentrations in such a way that the resistivity increases with high values of N_I.

With sample No. HA1 a value of $\mathscr{E}_A = 0{\cdot}22$ ev has been found in agreement with Hall, but with Nos. VA1 and VA7 the value obtained was 0·14 ev. Optical measurements by Burstein, Davisson, Bell, Turner and Lipson (1954) on indium-doped and zinc-doped germanium also give an activation energy of 0·14 ev. Measurements of Hall coefficient on gold-doped germanium by Dunlap (1953) give an energy level of 0·15 ev.

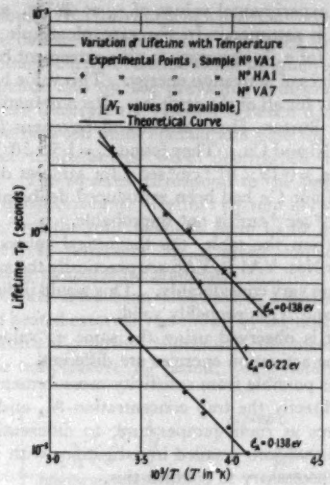


Figure 3.

With samples Nos. VU1 and VU2 no appreciable change in lifetime with temperature was detected below 290°K where the samples are in the extrinsic range of conductivity. This indicates that the activation energy is extremely

Activation Energy of Traps from Resistivity and Lifetime Measurements at 291°K and from the Temperature Dependence of Lifetimes

Sample	ρ (ohm cm)†	τ_p (μ sec)†	n_0+p_0 (10^{14} cm^{-3})†	N_I (10^{14} cm^{-3})	N_{ta} (10^{14} cm^{-3})	p_1 (10^{14} cm^{-3})†	$\mathscr{E}_{A\,(theor)}$ (ev)	$\mathscr{E}_{A\,(exp)}$ (ev)
VA1	8·3	35	2·4	30	13·8	947	0·138	0·138
VA2	16	70	1·3	30	14·4	1070	0·135	
VA3	4·3	40	4·6	15	5·2	782	0·142	
VA4	22	370	0·8	10	4·6	1140	0·133	
VA5	16	800	1·1	3	0·95	683	0·146	
VA6	15	700	1·3	3	0·95	632	0·148	
VA7	10	225	1·7	—	—	—	—	0·138
VU1	100	180	0·28	3000	1500	31600	0·035	very small
VU2	115	350	0·28	1000	500	40000	0·045	very small
HA1	10·6	180	1·6	$\leqslant 2$	$\leqslant 0{\cdot}2$	$\leqslant 47$	$\geqslant 0{\cdot}214$	0·22
HA2	10·6	300	1·6	$\leqslant 2$	$\leqslant 0{\cdot}2$	$\leqslant 78$	$\geqslant 0{\cdot}200$	
HA3	16	250	1·1	$\leqslant 2$	$\leqslant 0{\cdot}45$	$\leqslant 100$	$\geqslant 0{\cdot}214$	

† at 291°K.

In designation of samples, V means grown vertically, H grown horizontally, A grown from antimony-doped germanium, U from undoped germanium.

small. At temperatures above 290°K (intrinsic range) the lifetime becomes shorter. This is in agreement with equation (10).

Substituting appropriate values from the table in equation (5), it is found that the approximations used in the derivation of equation (6) are valid over a considerable range of temperatures.

A theoretical activation energy $\mathscr{E}_A$ of the traps can be calculated from equation (10), using the experimental values of τ_p at 291°K, n_0 and N_I, provided a suitable value for α_n is assumed. In the case of sample No. VA1 a value of $8{\cdot}17 \times 10^{-9}\,\text{sec}^{-1}\,\text{cm}^3$ for α_n is necessary to give agreement between the theoretical and experimental values of activation energy. This value has been used in evaluating $\mathscr{E}_A$ theoretically for all of the samples. Recombination constants have also been determined by Burton, Hull, Morin and Severiens (1953) on germanium samples doped with Ni and Cu. They found $\alpha_n = 1{\cdot}5 \times 10^{-9}\,\text{cm}^3\,\text{sec}^{-1}$ for samples doped with Ni and $\alpha_n = 0{\cdot}19 \times 10^{-9}\,\text{cm}^3\,\text{sec}^{-1}$ for samples doped with Cu.

Since neither Ni nor Cu had been introduced deliberately into our samples, the value $\alpha_n = 8 \times 10^{-9}\,\text{sec}^{-1}\,\text{cm}^3$ is not improbable.

As can be seen from the table, the theoretical values of activation energy obtained for samples Nos. VA1 to VA6 are practically the same, 0·14 ev, although the τ_p, n_0 and N_I values vary considerably. This would indicate that the predicted relation between τ_p and N_I is probably valid.

No disagreement is observed using the same α_n value with the remaining samples, although the activation energies are different.

It therefore seems possible from resistivity measurements at various temperatures to determine directly the trap concentration N_{ta} and, in conjunction with lifetime measurements at room temperature, to differentiate between samples having different traps. More detailed investigations with materials having high N_I values would be necesssary to confirm this.

ACKNOWLEDGMENT

Our thanks are due to Mr. L. J. Davies, Director of Research, for permission to publish this work.

REFERENCES

BURSTEIN, E., DAVISSON, J. W., BELL, E. E., TURNER, W. J., and LIPSON, H. G., 1954, *Phys. Rev.*, **93**, 65.

BURTON, J. A., HULL, G. W., MORIN, F. G., and SEVERIENS, J. C., 1953, *J. Phys. Chem.*, **57**, 853.

DUNLAP, W. C., Jr., 1953, *Phys. Rev.*, **91**, 1282.

HALL, R. N., 1952, *Phys. Rev.*, **87**, 835.

ROSE, F. W. G., and TIMMINS, E. W., 1953, *Proc. Phys. Soc.* B, **66**, 984.

SHOCKLEY, W., and READ, W. T., 1952, *Phys. Rev.*, **87**, 835.

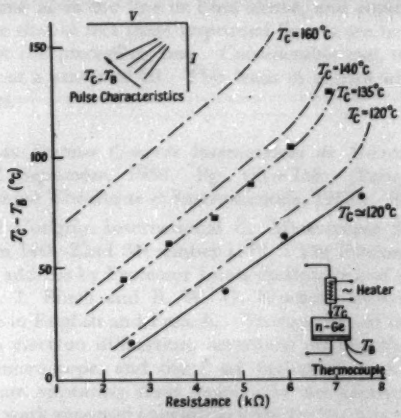
VALDES, L. B., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 1420.

LETTERS TO THE EDITOR

A Note on the Effect of Temperature Gradients at Point Contacts on Germanium

From the results obtained in the course of a previous investigation (Tipple and Henisch 1953) on the contact temperature of rectifying diodes, it was inferred that the isothermal voltage–current relation is not uniquely defined by any single temperature, but depends also on the prevailing temperature gradient. The chain lines on the figure illustrate the relation between d.c. resistance and total temperature difference between point contact and base for a typical case. The four different contact temperatures were determined thermoelectrically as previously described. The squares correspond to the peak back condition for which the current is almost constant. Neither current nor voltage is constant for the other chain lines, but the four relations seem to be members of the same family. The d.c. reverse resistance for a given contact temperature thus increases with increasing temperature gradient.

These conclusions have since been subjected to a further simple test. The whisker of a diode was heated externally as shown by the lower inset on the figure. The contact temperature T_C, again determined thermoelectrically, was kept constant and the base temperature T_B varied. The *isothermal* voltage–current



Effect of temperature gradients on the resistance of point contacts on n-type germanium.

—·—·— Inferred from d.c. measurements and the contact temperatures.

——— Pulse measurements.

relations of the diode were then determined as a function of $T_C - T_B$, using 1 μsec pulses. These relations are almost linear (see upper inset). The resistances obtained from their slopes have been entered on the figure as dots and are in general agreement with the results obtained by the previous method.

A possible mechanism of this effect has already been briefly discussed. If there is any appreciable hole current at large reverse voltages, it is reasonable to expect that the current should be sensitive to changes in the hole replacement rate. This should depend on the local temperature just inside the germanium, and hence on the temperature gradient, as well as on the immediate contact temperature. A similar effect could, in principle, arise from quite a different mechanism, e.g. in a single carrier system, if the voltage–current relation is primarily determined by a temperature-dependent spreading resistance. A numerical integration was carried out to estimate the importance of this effect for germanium of the particular resistivity and temperature dependence used. It showed that this mechanism would account for only a few per cent of the observed resistance changes.

The authors wish to thank Professor R. W. Ditchburn for his support, Dr. P. C. Banbury for helpful discussions and the Admiralty for permission to publish this note.

Department of Physics,
University of Reading.
15th June 1954.

H. K. Henisch.
P. M. Tipple.

Tipple, P. M., and Henisch, H. K., 1953, *Proc. Phys. Soc.* B, **66**, 826.

REVIEWS OF BOOKS

Four-Place Tables of Transcendental Functions, by W. FLÜGGE. Pp. 136. (London: Pergamon Press, 1954.) Price 25*s*.

Professor Flügge has prepared this book with a view to making the practical use of some of the transcendental functions easier, and therefore commoner. In addition to the trigonometric and hyperbolic functions, he has chosen to include the exponential function and its inverse, the logarithm; the two kinds, J and Y, of Bessel functions, of zero and first order, and their representatives for imaginary argument, I and K; the functions ber, bei, ker and kei, which give the real and imaginary parts of $J(x\sqrt{i})$ and $K(x\sqrt{i})$; the first derivatives of the last four; the elliptic integrals of the first and second kind; the error function; Fresnel's integrals; the sine, cosine and exponential integrals, of which the last is $\int_{-\infty}^{x} (e^{\beta}/\beta)\, d\beta$; and finally, the gamma function.

The tables are in nearly all cases to four significant figures, and great care has been taken to ensure that interpolation is simple; in every case, means of calculating values outside the range of tabulated arguments are given. There are also collections of formulae.

The lay-out of the tables seems to the present reviewer to have aimed at too much economy in printing. The first figures of an entry are not printed if they are the same as in the line or lines above, and copious use is made of asterisks to denote that in fact these unprinted figures are to be taken from the following, and not the preceding, line. Considerable use, too, is made of the device of taking out a power of 10. This leads to entries which are sometimes difficult to unravel.

J. H. AWBERY.

Comptes Rendus du Premier Congrès International de Microscopie Électronique, Paris, 14–22 *September*, 1950. Pp. viii + 768. (Paris: Editions de la Revue d'Optique Théorique et Instrumentale, 1953.) 8000 fr.

"Le Premier Congrès International de Microscopie Électronique" was held in Paris from 14th–22nd September 1950. The Proceedings, commencing with the opening address by Professor Louis de Broglie and inaugural addresses by Professors G. I. Finch and R. W. G. Wyckoff, include ninety-six communicated papers in English and French. Thirty-three of these are on electron optics, eleven on electron diffraction, seventeen on metallurgical applications of the electron microscope, and thirty on biological applications. Although the Proceedings are appearing three years after the Conference, a substantial proportion of the work reported appears in print for the first time.

The Conference was the first on electron microscopy at which most of the important schools in the field were represented. In particular, a strong contingent was present from Germany, the birthplace of the electron microscope. The Conference was held at a time when worldwide interest in the electron microscope was at its height, and the instrument was beginning to make progress —particularly in biological applications. A number of important new ideas were discussed for the first time at the Conference. These included the two possible methods, due to Gabor and Scherzer respectively, for improving the

resolving power beyond the limit set by the spherical aberration of the normal type of lens.

It is impossible to review adequately the wide range of papers included. They undoubtedly comprise the largest collection of papers concerned with recent developments in this field and, as such, take an important place in the literature. At the same time the papers have clearly not been refereed, and might have been improved by more strict editing.

The paper covers of the volume are a little inadequate to support the large number of pages, and it is aggravating to have to cut 120 pages before the papers can all be perused. The printing and diagrams are clear, but many of the reproductions of micrographs do not do full justice to the originals.

M. E. HAINE.

Soft Magnetic Materials for Telecommunications, edited by C. E. RICHARDS and A. C. LYNCH. Pp. viii + 346 (London: Pergamon Press, 1953). 63*s*.

In April 1952 a conference on soft magnetic materials and their applications, particularly for telecommunication purposes, was held at the Post Office Research Station, Dollis Hill, London. Some thirty-five papers were read, and these, in slightly abbreviated versions, together with some of the discussion they provoked, have now been collected and published in book form. The speakers were drawn from Universities and research establishments both from this country and the Continent, and as a result a large number of different aspects of soft magnetic materials are presented in this book by some of the best known authorities. Super specialists should perhaps be warned that the title has been interpreted rather liberally since the book contains two papers dealing with materials having coercivities of several hundreds! The bias throughout is very much towards the technical, and the reader is assumed to possess a quite extensive knowledge of ferromagnetic phenomena. One cannot help feeling that the value of the book would have been much increased and its public enlarged by the inclusion of an introduction mentioning the special nature of high permeability alloys and the proposed explanations of high permeability.

As it is, the book may be roughly divided into two sections, the first of which deals with various aspects of technical magnetization. This includes experimental and theoretical studies of losses under alternating field conditions (about 10 papers), time effect in magnetization (4), the study of surface layers on high permeability materials (2), and results of investigations on rectangular loop materials (2). The latter part of the book is largely devoted to magnetic materials (about 9). To the present reviewer the former seems to be the more stimulating and the one more worthy of preservation in book form, although there is much of interest in the latter. Unfortunately within this broad division there is so little coherence between the various papers that the whole volume reads more like a year's issue of the journal of a society for soft magnetic materials than a book in the customary sense of the word.

It is sometimes felt that academic physicists know all too little about the problems of interest to their colleagues in technical research centres. This book should help to provide those who so desire with one way of overcoming their ignorance—particularly those seeking new problems, for there are plenty in this field. To those, and the specialists in soft magnetic materials, this book may be recommended. It is very well produced and remarkably free from typographical errors.

E. W. LEE.

The Determination of Crystal Structures, by H. LIPSON and W. COCHRAN. Pp. x+345. (London: G. Bell and Sons Ltd., 1953.) 50*s*.

This is the third volume of a series which began in 1933 with the publication of *The Crystalline State: A General Survey*, by W. L. (now Sir Lawrence) Bragg, and continued with *The Optical Principles of the Diffraction of X-Rays*, by R. W. James (1948). The three volumes differ considerably in difficulty, and the third is related more closely in spirit and topic with the first than with the second. It takes up the problem of crystal-structure determination at the point at which a set of experimental structure amplitudes is available, and describes first the techniques that can be applied as a matter of routine, then those that may lead to the finding of a trial structure, then methods of refinement, and finally methods of estimating the accuracy of the refined structure.

The subject is one in which there have been great developments during the past few years, and the way in which the authors have digested the new techniques and exhibited the unity underlying the apparent diversity of different approaches is most remarkable, though not surprising to those acquainted with the senior author's expository papers. This will be a great aid to the established crystallographer and an even greater one to the beginner. Equally remarkable is the completeness with which the subject is covered; not only are the techniques of established usefulness included, but also suggestions of undeveloped promise. [One of the latter that does appear to have been ovlerooked is the method of Patterson interpretation suggested by Sanadze and Ždanov in *Dokl. Akad. Nauk S.S.S.R.*, 1950, **73**, 111. This is based on the fact that those light-atom–heavy-atom peaks that are centro-symmetrically related about the centre of a heavy-atom–heavy-atom vector form an image of the crystal structure, possibly complicated by ghosts of various types.] The literature is well covered up to 1952, and there are a few 1953 references.

The book contains a number of innovations in methods of cross reference. To explain these in detail would require more space than can be spared in a review; it will suffice to say that all work smoothly within a chapter, and that between chapters they are vitiated to some extent by the absence of an indication of the chapter number on each page—a useful feature of several recent books. One innovation for which no good reason is apparent is that upper-case symbols are printed in roman type, although lower-case symbols appear in the usual italic. No serious misprints have been noted; it may be mentioned that the angle marked θ in figure 5 (*b*) is the complement of the Bragg angle, cyanite is triclinic on page 112 but cubic on page 120, and Menke has a supernumerary '*c*' throughout.

All concerned with crystal-structure determination are greatly indebted to the authors and publishers of this series. They will not take it as unfair criticism if I end by pleading for another volume: Experimental Methods. There appears to be no book in English giving really practical descriptions of the photographic and other methods and apparatus in use for recording and measuring the intensities of x-ray diffraction maxima. Such a book would greatly ease the task of those whose task it is to teach (or learn) x-ray crystallography.

A. J. C. WILSON.

CONTENTS OF SECTION A

PAGE

Miss M. H. Alston, Dr. A. V. Crewe, Dr. W. H. Evans, Dr. L. L. Green and Dr. J. C. Willmott. The Scattering of 15·7 Mev Neutrons by ^{4}He . . 657

Prof. D. R. Bates and Major G. W. Griffing. Inelastic Collisions between Heavy Particles: II—Contributions of Double-Transitions to the Cross Sections associated with the Excitation of Hydrogen Atoms in Fast Encounters with other Hydrogen Atoms 663

Mr. K. Phillips. On the Thick Target Bremsstrahlung Spectrum at Relativistic Energies 669

Mr. W. Rothenstein. The Second Born Approximation in Inelastic Collisions of Electrons with Atoms 673

Mr. N. T. S. Evans and Dr. W. C. Parkinson. Angular Distributions in the ^{10}B(d, p)^{11}B Reaction 684

Prof. H. S. W. Massey and Prof. C. B. O. Mohr. Gaseous Reactions involving Positronium 695

Mr. K. M. Howell. The van der Waals Energy of Two Helium Atoms . . 705

Prof. C. B. O. Mohr and Mr. L. J. Tassie. The Scattering and Polarization of Electrons by Gold 711

Mr. I. E. McCarthy and Prof. H. S. Green. A Method for the Solution of Nuclear Bound-State Problems 719

Letters to the Editor:

Dr. N. L. Balazs. On the Propagation of Energy in Elastic Media . . 726

Mr. W. R. G. Kemp, Dr. P. G. Klemens, Mr. A. K. Sreedhar and Dr. G. K. White. The Lattice Thermal Conductivity of Silver-Palladium Alloys at Low Temperatures 728

Prof. C. B. O. Mohr. The Difference in the Multiple Scattering of Electrons and Positrons 730

Prof. W. E. Burcham. The α-Activity induced in Gold by Bombardment with Ions of Carbon 13 733

Dr. G. S. Bogle and Mr. V. Heine. Paramagnetic Resonance in Gadolinium Sulphate Octohydrate 734

Reviews of Books 737

Contents of Section B 740

Fluorescence Spectra of Organic Crystals

By J. B. BIRKS† and G. T. WRIGHT
Physics Department, Rhodes University, Grahamstown, South Africa

MS. received 6th April 1954, and in amended form 14th June 1954

Abstract. Relative quantum intensity measurements have been made of the fluorescence spectra, excited by 254 mμ radiation, of crystalline anthracene, *trans*-stilbene, *para*-terphenyl and diphenylacetylene at room temperature, using a calibrated spectrophotometer and photomultiplier. Observations were made of the reflection and transmission spectra from 1 cm thick crystals of each compound, and also of the reflection spectrum from a thin microcrystalline layer of anthracene. The latter corresponds closely to the true molecular emission. Due to self-absorption, this has only a probability $k_e = 0{\cdot}24$ of direct transmission through a thick crystal. This value confirms the difference in technical and molecular photo-fluorescence decay times, reported by Birks and Little (1953). For the other materials, k_e is larger, and for *para*-terphenyl it is practically unity.

Vibrational bands have been resolved, which have been assigned to the 1–0 and 0–n ($n = 0$ to 3) transitions in anthracene, and to the 0–n ($n = 0$ to 5) transitions in *para*-terphenyl. The 1–0 transition in anthracene is attributed to fluorescence prior to the establishment of thermal equilibrium.

§ 1. Introduction

Anthracene, *trans*-stilbene, *para*-terphenyl and diphenylacetylene are among the most efficient organic crystalline phosphors known and they are hence of particular interest for scintillation counting (Birks 1953). Many of the fluorescence characteristics of these crystals have been studied in detail. These characteristics include the scintillation and photo-fluorescence decay times and their dependence on temperature, the scintillation efficiency and its temperature dependence, the scintillation response to ionizing particles of different specific ionization and of different range, and the photo-fluorescence quantum efficiency. A quantitative theory of the scintillation and fluorescence properties of organic crystals and solutions has been developed which is in general agreement with these and other experimental data (Birks 1953, 1954).

Observations have also been made of the fluorescence and absorption spectra of some of these compounds in liquid solution. Data on the spectra of the crystalline materials at room temperature are however meagre, apart from the measurements of Kortum and Finckh (1942) and Bowen and Lawley (1949) on anthracene. Koski (1951), Koski and Thomas (1951) and Sangster (1952) have recorded fluorescence spectrograms for each of the four compounds, but with arbitrary intensity scales, and for thick crystals only.

Birks and Little (1953) have compared the fluorescence spectra of microcrystalline layers and thick crystal specimens of anthracene, using a calibrated photomultiplier and spectrophotometer. These measurements confirm those of Kortum and Finckh (1942) and show that, due to the overlap of the fluorescence

† Now at British Dielectric Research, 38 Wood Lane, London W.12.

and absorption spectra, there is a major difference between the molecular (microcrystalline) emission spectrum and the technical emission spectrum, observed in transmission through a thick crystal. Approximately 80% of the fluorescence emission from the anthracene molecules is reabsorbed within the crystal. This self-absorption, which is followed by further emission and absorption processes, increases the molecular photo-fluorescence decay time from $3{\cdot}5 \pm 1{\cdot}0$ mμsec to its technical (thick crystal) value of 14 ± 2 mμsec. This phenomenon is part of the general photon cascade process described by Birks (1953, 1954).

The spectral resolution obtained by these authors was limited by the low intensity of the monochromatic light source available, and the vibrational structure of the electronic transition was not resolved. These measurements have therefore been repeated with greatly improved resolution, the fluorescence spectrum has also been observed in reflection, and the observations have been extended to include *trans*-stilbene, *para*-terphenyl and diphenylacetylene. All measurements were made at room temperature.

§2. Experimental Method and Results

Monochromatic radiation of 253·7 mμ wavelength was used for the excitation of the fluorescence spectra. The extinction coefficient of each of the crystals studied is so high at this wavelength that the exciting radiation is completely absorbed in a thickness of the order of a wavelength or less. An intense steady radiation source was obtained using a 7-watt Model 11A Hanovia mercury discharge tube, in which a large fraction of the radiation energy is of 253·7 mμ wavelength. The thin discharge tube was placed centrally in the entrance plane of a medium quartz spectrograph ($F_D = 60$ cm) from which the normal entrance slit mechanism had been removed. An intense image of the 253·7 mμ line, completely resolved from the other mercury emission lines, was formed in the exit plane of the spectrograph. The radiation was then focused on to the specimen under investigation.

The fluorescence spectra were observed with a Cenco-Sheard 'spectrophotelometer', using an EMI 6262 photomultiplier tube, operating from a stabilized power unit and coupled to a sensitive galvanometer, as a detector. This spectrometer has a constant wavelength band width from the nature of the grating mounting employed, and the variation of its transmission coefficient with wavelength has been previously measured (Birks and Szendrei 1953). The relative spectral response of the photomultiplier has also been measured, using the calibrated spectrometer and a standard lamp of known colour temperature, supplied by the National Physical Laboratory, Teddington. The readings of the photomultiplier output current were therefore corrected for the overall spectral response of the spectrometer and photomultiplier to give readings proportional to the intensity I of the fluorescence. The small, steady dark current from the photomultiplier was balanced out. In plotting the spectra the relative quantum intensity, proportional to $I/h\nu$, is plotted as a function of the wavelength λ.

The fluorescence spectrum of anthracene was observed under four conditions: (*a*) in reflection from a very thin microcrystalline layer evaporated from xylene solution on to a quartz plate, (*b*) in reflection from a 1 cm thick single crystal, (*c*) in transmission through the 1 cm thick single crystal, and (*d*) in transmission

through a 1 mm thick crystal of naphthalene containing a dilute solid solution of anthracene.

Spectra (*a*) and (*b*) were normalized at a wavelength of 460 mμ and were found to coincide at longer wavelengths. Spectra (*b*) and (*c*) were normalized at a wavelength of 445 mμ and were found to coincide at longer wavelengths. The spectra (*a*), (*b*) and (*c*) of anthracene are plotted in figure 1.

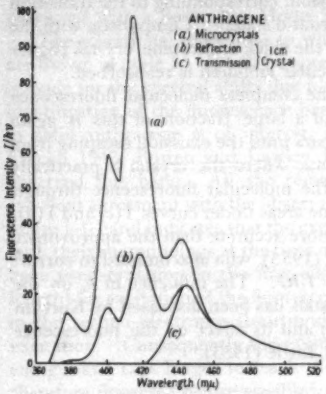


Figure 1. Anthracene fluorescence.

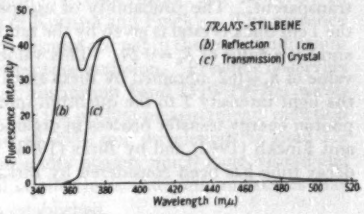


Figure 2. *Trans*-stilbene fluorescence.

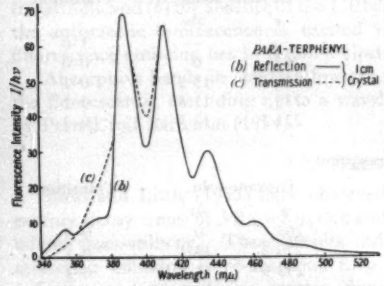


Figure 3. *Para*-terphenyl fluorescence.

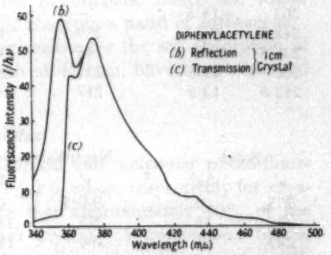


Figure 4. Diphenylacetylene fluorescence.

The fluorescence spectra of *trans*-stilbene, *para*-terphenyl and diphenylacetylene were measured under conditions (*b*) and (*c*) above. These measurements showed that the self-absorption in these materials was much less than in anthracene, and hence it was not necessary to use condition (*a*) to resolve the vibrational structure of the molecular fluorescence bands. A similar normalization procedure was followed. The results are plotted in figures 2, 3 and 4. The single pure crystals used were supplied by National Radiac Inc., and were of high quality and clarity. We are indebted to Professor J. W. Cook for the gift

of the pure anthracene used for spectra 1 (*a*) and (*d*). The naphthalene used in (*d*) was purified by successive crystallization.

§ 3. Discussion

3.1. *Anthracene*

The microcrystalline fluorescence spectrum of anthracene (figure 1 (*a*)) approximates to the complete molecular emission, corresponding to the transition from the first excited electronic state to the ground state.† Comparison with the spectrum observed in transmission through the thick anthracene crystal (figure 1 (*c*)) shows that a large fraction of the molecular emission is reabsorbed.

The absorbed radiation then re-excites the complete molecular fluorescence with a quantum efficiency of about 0·9, and a large fraction of this is again reabsorbed. This photon cascade process recurs until the emission escaping from the crystal lies at wavelengths beyond 420 mμ where the crystal is practically transparent. The probability of escape of the molecular fluorescence through the 1 cm thick crystal is given by the ratio of the areas under curves 1 (*c*) and 1 (*a*), and has a value of $k_e = 0{\cdot}24$. This value is more accurate than the approximate value of $k_e = 0{\cdot}2$, obtained by Birks and Little (1953), who also omitted to correct the light intensity I to the quantum intensity $I/h\nu$. The influence of k_e on the photon energy transfer process in organic crystals has been discussed by Kortum and Finckh (1942) and by Birks (1953, 1954) and its effect on the fluorescence decay time has been considered by Birks and Little (1953).

Table 1. Anthracene

(*a*) Fluorescence

Solid		Naphthalene soln.		Dioxane soln.		Transition
ν'	$\Delta\nu'$	ν'	$\Delta\nu'$	ν'	$\Delta\nu'$	
263	−14·5	—	—	—	—	1–0
248·5	0	259·5	0	262	0	0–0
239·5	9	245	14·5	248·5	13·5	0–1
226	13·5	231	14	235	13·5	0–2
212·5	13·5	217	14	221·5	13·5	0–3

(*b*) Absorption

Solid		Naphthalene soln.		Dioxane soln.		Transition
ν'	$\Delta\nu'$	ν'	$\Delta\nu'$	ν'	$\Delta\nu'$	
255	0	260·5	0	265	0	0–0
270	15	275	14·5	279·3	14·3	0–1
285	15	290	15	293·6	14·3	0–2
299	14	—	—	308	14·4	0–3

The ratio of the areas under curves 1 (*b*) and 1 (*a*) is 0·456, indicating that only this fraction of the molecular emission immediately escapes from the front face of the thick crystal. All the vibrational bands of the molecular fluorescence spectrum are, however, observed in the reflection spectrum (figure 1 (*b*)).

The wave-numbers ν' and wave-number intervals $\Delta\nu'$ (each in $10^2\,\mathrm{cm}^{-1}$) of the maxima of the vibrational bands resolved in the fluorescence spectrum of microcrystalline anthracene (figure 1 (*a*)) are listed in table 1 (*a*). Similar data

† The predicted fluorescence due to transitions from the higher excited electronic states (Birks 1953, 1954) is not within the spectral range studied, and is the subject of another investigation now in progress.

are also given for the fluorescence spectrum of anthracene, (i) in dilute solid solution in naphthalene, and (ii) in solution in dioxane (Kortum and Finckh 1942).

Table 1 (*b*) lists the values of ν' and $\Delta\nu'$ for the principal maxima in the corresponding absorption spectrum of anthracene, (i) for solid crystals (Kortum and Finckh 1942, Seshan 1936), (ii) in dilute solid solution in naphthalene, and (iii) in solution in dioxane (Kortum and Finckh 1942).

The absorption spectrum of the anthracene in naphthalene solution was measured with a Beckmann Model DU spectrophotometer.

The vibrational transitions to which the individual bands have been assigned are shown in table 1. It will be noted that all the data are consistent, if allowance is made for the spectral shifts due to the different molecular environments. The identification of the fluorescence band at about 26 300 cm^{-1} as the 1–0 transition in solid anthracene is of interest, since previous observers (e.g. Kortum and Finckh 1942, Bowen and Lawley 1949, Pesteil and Barbaron 1954) have not observed this band. All the remaining fluorescence bands are, however, in excellent agreement with the observations of Kortum and Finckh. The presence of the 1–0 band indicates that the excited molecules are not in thermal equilibrium with their surroundings at the time of fluorescence emission. Previous observers have used radiation in the first electronic absorption band (310–390 mμ wavelength) for excitation of the fluorescence. In the present measurements radiation in the second electronic absorption band of 253·7 mμ wavelength is used for excitation. Consequently the excited molecules possess much more vibrational energy, and take longer to reach thermal equilibrium, and some molecules may therefore fluoresce before equilibrium is established.

This interpretation is supported by (*a*) the agreement of $\Delta\nu'$ between the 1–0 and 0–0 fluorescence transitions and $\Delta\nu'$ between the 0–0 and 1–0 absorption transitions, and (*b*) the absence of the 1–0 band in the naphthalene solution, in which the anthracene fluorescence is excited via the naphthalene molecules, whose fluorescence emission lies in the first electronic absorption band of anthracene.

Absorption bands in solid anthracene, responsible for the self-absorption of the fluorescence, extending up to a wavelength of 420 mμ, have been observed by Pesteil and Barbaron (1954).

3.2. *Trans-Stilbene*

Birks and Little (1953) have observed technical and molecular photo-fluorescence decay times of $3{\cdot}0 \pm 0{\cdot}8$ mμsec and $1{\cdot}7 \pm 0{\cdot}6$ mμsec respectively for crystalline *trans*-stilbene. These results indicate that approximately 40% of the molecular emission is re-absorbed in a thick crystal. A comparison of the reflection and transmission spectra (figure 2) confirms the existence of a self-absorption of this order of magnitude.

Table 2. *Trans*-Stilbene (Fluorescence)

ν'	277	261	245	230	~216
$\Delta\nu'$	0	16	16	15	~14

The wave-numbers and wave-number intervals (in 10^2 cm^{-1}) of the maxima of the vibrational bands resolved in the reflection fluorescence spectrum (figure 2 (*b*)) are listed in table 2. The measurements do not extend to sufficiently short wavelengths to identify the transitions with certainty. The values of $\Delta\nu'$ are in approximate agreement with those of Hausser *et al.* (1935), who have reported

that at room temperature the maxima in the fluorescence spectra of the diphenylpolyenes, of which *trans*-stilbene is the first member, are equidistant with nearly the same constant spacing $\Delta\nu' = 1430\,\text{cm}^{-1}$. The same authors have reported that the spectra at low temperatures can be represented by the equation

$$\nu' = \nu_0' - 1550n_1 - 1160n_2\,\text{cm}^{-1},$$

with $n_1 = 0, 1, 2, 3, 4$ and $n_2 = 0, 1, 2$, and $\nu_0' = 29\,730$ for *trans*-stilbene. As Pringsheim (1949) has pointed out, if this representation is correct, a spacing of $\Delta\nu'$ of $1430\,\text{cm}^{-1}$ at room temperature can only be fortuitous and caused by the superposition of the two oscillations of $1550\,\text{cm}^{-1}$ and $1160\,\text{cm}^{-1}$. Moreover, there is doubt about the identification of ν_0' with the 0–0 transition.

The absorption responsible for the difference between the transmission and reflection spectra (figure 2) has been observed by Pesteil (1951), who has studied its polarization characteristics. The main absorption peaks are at 387 mμ and 367 mμ. The absorption spectrum in solution reported by Friedel and Orchin (1951) shows no vibrational structure.

3.3. *Para-Terphenyl*

The fluorescence spectra of *para*-terphenyl (figure 3) are of interest since they indicate that self-absorption is negligible in this material. This agrees with the observations of Birks and Little (1953), who found that the photo-fluorescence decay times of a thick crystal and a fine powder were identical, within the experimental error. The two spectra are identical, apart from slight differences at wavelengths between the two principal maxima, and below 380 mμ. The measurements below 360 mμ indicate the possible existence of a very weak absorption band in this region.

Table 3. *Para*-Terphenyl (Fluorescence)

Crystal		Solution in xylene		Transition
ν'	$\Delta\nu'$	ν'	$\Delta\nu'$	
282	0	290	0	0–0
270	12	274	16	0–1
258	12	258	16	0–2
244	14	—	—	0–3
230	14	—	—	0–4
217	13	—	—	0–5

The wave-number ν' and wave-number intervals $\Delta\nu'$ (each in $10^2\,\text{cm}^{-1}$) of the maxima observed in the reflection fluorescence spectrum are listed in table 3. Similar data for the fluorescence spectrum of *para*-terphenyl in xylene solution (Reynolds 1952) are shown for comparison. Unfortunately the solution absorption spectrum recorded by Friedel and Orchin (1951) shows no vibrational structure, and the measurements do not extend to wavelengths above 320 mμ, where the absorption is still high ($\log_{10}\epsilon = 3{\cdot}6$, at 320 m$\mu$). In the absence of absorption data, the identification of the 0–0 transition can only be tentative, but the transition assignments shown in table 3 are suggested.

3.4. *Diphenylacetylene*

The transmission and reflection fluorescence spectra of diphenylacetylene (figure 4) show that the self-absorption in this material is relatively small. This is consistent with the short decay time of $3{\cdot}0 \pm 0{\cdot}8$ mμsec observed for thick crystals

(Birks and Little 1953). The diffuse spectrum between 380 mμ and 420 mμ is probably due to the superposition of two overlapping vibrational bands with maxima at about 392 mμ and 410 mμ. The values of ν' and $\Delta\nu'$ (in $10^2\,cm^{-1}$) of the vibrational band maxima observed, including the two interpolated maxima, are listed in table 4. Similar data for the absorption spectrum in cyclohexane solution (Friedel and Orchin 1951) are also shown. The interval beween the fluorescence and absorption spectra is too large to attempt a positive identification of the vibrational transitions.

Table 4. Diphenylacetylene

Crystal fluorescence		Absorption in cyclohexane	
ν'	$\Delta\nu'$	ν'	$\Delta\nu'$
279	0	335·5	0
266·5	12·5	346	10·5
~255	~11·5	356·5	10·5
~244	~11	366	9·5
~233	~11	377	11

§4. Conclusion

The fluorescence spectra of the four organic crystals studied indicate the importance of allowing for any overlap with the absorption spectra. The observations are consistent with the decay time measurements of Birks and Little (1953), and confirm the self-absorption effects reported by them. It has been found possible to analyse the vibrational structure of the spectra, and in some cases to assign the bands observed to various energy transitions. These transitions show certain regularities and similarities in the four compounds studied. The experimental method is applicable to other fluorescence compounds, and should provide useful data on the molecular vibrational levels in the crystalline state for comparison with the Raman spectra.

Acknowledgments

We are indebted to the University of Stellenbosch for the load of the medium quartz spectrograph, and to the South African Council for Scientific and Industrial Research for a grant in support of this work. We wish to thank Mr. F. S. Allison for the meticulous care taken in recording the spectra.

References

Birks, J. B., 1953, *Scintillation Counters* (London : Pergamon Press; New York : McGraw–Hill); 1954, *Phys. Rev.*, **94**, 1567.

Birks, J. B., and Little, W. A., 1953, *Proc. Phys. Soc.* A, **66**, 921.

Birks, J. B., and Szendrei, M. E., 1953, *Phys. Rev.*, **91**, 197.

Bowen, E. J., and Lawley, P. D., 1949, *Nature, Lond.*, **164**, 572.

Friedel, R. A., and Orchin, M., 1951, *Ultra-Violet Spectra of Aromatic Compounds* (New York : John Wiley; London : Chapman and Hall).

Hausser, K. W., Kuhn, R., and Kuhn, E., 1935, *Z. Phys. Chem.* B, **29**, 417.

Kortum, G., and Finckh, B., 1942, *Z. Phys. Chem.* B, **52**, 263.

Koski, W. S., 1951, *Phys. Rev.*, **82**, 230.

Koski, W. S., and Thomas, C. O., 1951, *J. Chem. Phys.*, **19**, 1286.

Pesteil, P., 1951, *C. R. Acad. Sci., Paris*, **233**, 924; 1952, *Ibid.*, **235**, 150.

Pesteil, P., and Barbaron, M., 1954, *J. Phys. Radium*, **15**, 92.

Pringsheim, P., 1949, *Fluorescence and Phosphorescence* (New York : Interscience).

Reynolds, G. T., 1952, *Nucleonics*, **10**, No. 7, 46.

Sangster, R. C., 1952, *M.I.T. Technical Report*, 55 (Jan. 1).

Seshan, P. K., 1936, *Proc. Indian Acad. Sci.* A, **3**, 148.

Heat Flow in the Transvaal and Orange Free State

By A. E. CARTE
National Physical Laboratory, South African Council for Scientific and Industrial Research, Pretoria, S. Africa

Communicated by S. M. Naudé; *MS. received 4th May* 1954

Abstract. The thermal conductivity of 46 rock samples and the temperature at a number of points along the lengths of three boreholes were measured. The heat flow from the earth is derived from these and other measurements of temperature and conductivity. Values of the heat flow lie in the region $(1{\cdot}1 \pm 0{\cdot}3) \times 10^{-6}$ cal cm^{-2} sec^{-1}.

§ 1. Introduction

In the course of the past five years the Heat Subdivision of the South African National Physical Laboratory has carried out measurements of temperature in boreholes and thermal constants of rocks. Descriptions of parts of this work together with some results have been published by Mossop (1950) and Mossop and Gafner (1951). Further results, including calculations of heat flow from the earth, are given here.

§ 2. Measurements

The apparatus used to measure the thermal conductivity of rock specimens was of the divided bar type (Mossop and Gafner 1951). All conductivities in this paper are expressed in g cal cm^{-1} sec^{-1} deg^{-1}. Temperatures in the boreholes were measured by the multi-thermometer method that has been described by Mossop (1950). The mercury maximum thermometers that were used were standardized to $\pm 0{\cdot}02$°C.

§ 3. Observations

3.1. *Borehole HB* 15

Borehole HB 15 is situated in the Klerksdorp District of the Transvaal. Drilling to a depth of 6586 ft was completed in August 1950 and the temperatures were measured three months later. The estimated accuracy of the temperatures is to $\pm 0{\cdot}05$°C. The thermal conductivity of samples from 32 different depths at 200 ft intervals was measured; mean values and standard errors are given in table 1.

Table 1. Thermal Conductivity of Rocks from HB 15

	Dolomite	Ventersdorp lava	Witwatersrand quartzite	Dyke
Mean conductivity $\bar{k}$	0·0132	0·0071	0·0128	0·0107
Standard error $\sigma(\bar{k})$	$9{\cdot}6 \times 10^{-4}$	$2{\cdot}3 \times 10^{-4}$	$5{\cdot}6 \times 10^{-4}$	—
No. of samples n	8	11	11	1

$\sigma(\bar{k}) = [\{\Sigma(k-\bar{k})^2\}/n(n-1)]^{1/2}$, where k is the measured conductivity and $\bar{k}$ the arithmetic mean.

The heat flow at depth D, where the temperature is T and conductivity k, is given by

$$H = k\,dT/dD. \qquad \ldots\ldots(1)$$

Thus, in a region where the temperature–depth curve is a straight line and measured conductivities show no systematic trend, it is reasonable to conclude that, in this region, (i) H is constant, (ii) k is constant, (iii) T measurements are reliable and undisturbed by factors such as water movement and climatic changes. When these conditions apply, equation (1) can be re-arranged as the linear equation

$$T = T_0 + HD/k \qquad \ldots\ldots(2)$$

where T_0 is a constant.

The temperature–depth curve for HB 15 (figure 1) shows three distinct

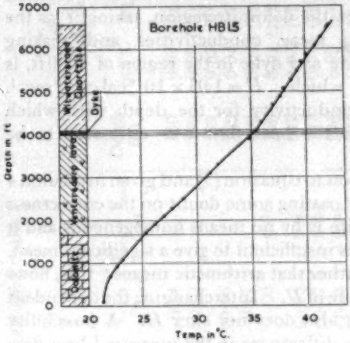


Figure 1.

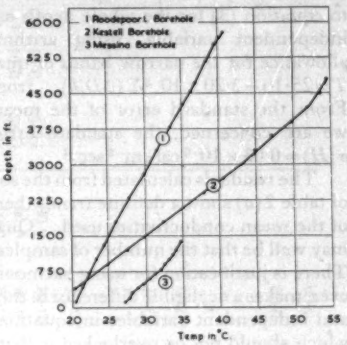


Figure 2.

sections corresponding to the three principal lithological sections. The temperature gradient in the dolomite is by no means constant, although if the first point is omitted the next four points do lie on a straight line. The solution to equation (2) found by fitting the least-squares straight line to these four points is

$$T = 19{\cdot}50 + 3{\cdot}44 \times 10^{-3} D$$

when D is expressed in feet. Multiplying the slope of this line by the arithmetic mean conductivity of the dolomite gives $H = 1{\cdot}49 \times 10^{-6}$ cal cm^{-2} sec^{-1}. Even though the differences, shown in column 3 of table 2 (*a*), between the observed temperatures and those calculated from the equation are very small, indicating that the points all lie close to a straight line, this value of H cannot be considered reliable because of (i) the small number of temperature observations, (ii) the large random variations of the measured conductivities, (iii) the conclusion reached by Bullard (1939) and Bouwer (1953) that water movement and other factors influence temperatures in dolomite, particularly near the surface.

In both the Ventersdorp lava and the Witwatersrand quartzite sections none of the points deviates far from the straight lines. The equations of these two straight lines are given below and the differences between the observed and calculated temperatures are shown in column 3 of table 2 (*a*). In the lava

$$T = 18{\cdot}07 + 4{\cdot}33 \times 10^{-3} D, \quad H = 1{\cdot}00 \times 10^{-6} \text{ cal cm}^{-2} \text{ sec}^{-1}.$$

If we assume that the temperature gradient is known precisely, then the standard error in H is $\sigma(H) = H\sigma(\bar{k})/\bar{k} = 0{\cdot}03 \times 10^{-6}$ cal cm^{-2} sec^{-1}. Similarly, for the quartzite,

$$T = 24{\cdot}06 + 2{\cdot}77 \times 10^{-3} D, \quad H = 1{\cdot}15 \times 10^{-6} \text{ cal cm}^{-2} \text{ sec}^{-1},$$
$$\sigma(H) = 0{\cdot}05 \times 10^{-6} \text{ cal cm}^{-2} \text{ sec}^{-1}.$$

Since the data for the lava and quartzite sections appear reliable and the heat flow values, separately calculated, show agreement to within the estimated accuracies, all the observations can be combined to give a single better value of H according to the method described by Bullard (1939). Equation (1) can be written in the form

$$T = T_0 + H\Sigma(\Delta D/k). \qquad \ldots\ldots(3)$$

This shows a linear relation between T and $\Sigma(\Delta D/k)$, where D is the depth of a thermally isotropic region of conductivity k. The least-squares solution to equation (3) for the whole depth except the dolomite region, taking T as the independent variable, using arithmetic mean conductivities and making allowance for the narrow bands of quartzite and dyke in the region of 4000 ft, is $T = 25{\cdot}10 + 3{\cdot}20 \times 10^{-5}\Sigma(\Delta D/k)$, from which $H = 1{\cdot}05 \times 10^{-6}$ cal cm^{-2} sec^{-1}. From the standard error of the mean conductivity for the depth with which we are concerned, the standard error of the heat flow was estimated to be $\sigma(H) = 0{\cdot}03 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

The residuals calculated from the solution to equation (3) and given in column 4 of table 2 (*a*) show a definite trend, thereby casting some doubt on the correctness of the mean conductivities used. Quartzite is by no means homogeneous and it may well be that the number of samples was insufficient to give a significant mean. There is justification for using harmonic rather that arithmetic means: this, however, makes a negligible difference to the value of H. Interchanging the dependent and independent variables in equation (3) also does not alter H. A possibility which should not be overlooked is that the differences in the measured heat flow in the lava and quartzite are real; climatic changes may have disturbed the temperature equilibrium or the heat flow in this region may be non-vertical.

To summarize, the observations in HB 15 lead to the value for the mean vertical heat flow of $H = 1{\cdot}05 \times 10^{-6}$ cal cm^{-2} sec^{-1} and $\sigma(H) = 0{\cdot}03 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

3.2. *Roodeport Borehole No.* 1

Roodeport borehole No. 1 (R 57) is in the Klerksdorp District of the Transvaal. Drilling to a depth of 5860 ft was completed in 1938. Eleven months after drilling had ceased the temperatures at 17 points were measured by the Geological Survey (Krige 1948). Twelve years later the author repeated the measurements at the same points. The results of this comparison are given in column 4 of table 2 (*b*). It will be seen that only in one case, viz. 100 ft, does the difference in observed temperatures exceed 0·1°C. For the 17 readings the mean difference is 0·02°C. The difference at 100 ft is too great to be accounted for by diurnal or seasonal temperature changes, but may be due to water movement. The conclusion to be drawn is that this borehole, which took eight months to drill, reached a state of temperature equilibrium within eleven months after drilling operations had ceased. Also the accuracy of the temperature measurements is confirmed.

Heat flow.

Column 3 of table 2 (*b*) shows that in the lava there is a significant decrease in geothermic step with increasing depth. Conductivity measurements of 15 samples of lava from this hole show no corresponding systematic trend, being scattered about a mean of 0·0072 (Mossop and Gafner 1951). This value is very

Table 2. Borehole Temperature Results

(a) HB 15

Rock	Depth (ft)	$T_{obs}-T_{calc}$ (°c) (eqn (2))	$T_{obs}-T_{calc}$ (°c) (eqn (3))
	100		
Dolomite	697	−0·01	
	896	+0·04	
	1096	−0·08	
	1693	+0·02	+0·20
	1892	+0·07	+0·25
	2092	+0·02	+0·15
Ventersdorp lava	2291	−0·01	+0·07
	2490	−0·08	−0·04
	2888	−0·08	−0·13
	3090	−0·01	−0·11
	3291	−0·02	−0·16
	3888	+0·09	−0·13
	4086		−0·23
	4285	+0·05	−0·26
	4685	+0·02	−0·19
	4882	+0·01	−0·14
	5084	−0·06	−0·16
Witwatersrand quartzite	5281	−0·06	+0·01
	5480	+0·02	+0·02
	5681	−0·16	−0·10
	5882	−0·15	−0·04
	6082	+0·02	+0·19
	6279	+0·04	+0·26
	6480	+0·12	+0·40
	6550	0·00	+0·29

(b) Roodepoort

Rock	Depth (ft)	Geothermic step (ft deg^{-1})	$T_{1938}-T_{1951}$	$T_{obs}-T_{calc}$ (°c)
Lava	100		+0·14	
	404	301	+0·02	
Agglomerate	705	317	+0·06	
	1006	276	−0·01	
	1305	290	0·00	−0·01
Lava	1801	286	+0·03	+0·02
	2298	296	+0·07	−0·04
	2794	271	+0·04	+0·01
	3290	271	+0·01	+0·02
	3786	271	0·00	−0·01
	4283	260	−0·06	0·00
	4555	259	−0·04	0·00
Impure argillaceous quartzite	4727	302	+0·01	
Quartzite	5177	300	−0·01	
Quartzite partly argillaceous and micaceous	5522	303	+0·06	
Lava	5683	218	0·00	
Quartzite	5783	250	−0·06	

(c) Messina

Depth (ft)	$\sigma(T)$	(n)	$T_{obs}-T_{calc}$ (°c) (eqn (3))
100	0·02	(6)	
199	0·01	(11)	
299	0·02	(6)	−0·01
398	0·01	(6)	−0·01
498	0·01	(6)	0·00
598	0·01	(12)	+0·05
697	0·02	(6)	+0·02
797	0·01	(9)	−0·04
896	0·03	(5)	
996	0·01	(8)	
1096	0·02	(6)	
1195	0·01	(7)	
1295	0·01	(4)	

(d) Kestell

Depth (ft)	$T_{obs}-T_{calc}$ (°c) (eqn (3))
200	−0·47
400	−0·16
600	−0·26
800	−0·16
1000	−0·05
1200	+0·25
1400	−0·05
1600	−0·18
1800	−0·14
2000	+0·28
2200	+0·49
2400	+0·89
2600	+1·17
2800	+1·00
3000	+0·92
3200	+0·21
3400	+0·29
3600	−0·08
3800	−0·41
4000	−0·27
4200	−0·53
4400	−1·17
4600	−1·51

close to the mean given in table 1 and to the value for Ventersdorp lava given by Mossop and Gafner (1951). Factors which might account for the decrease in geothermic step are: (i) The thermal conductivity of the lava *in situ* might have a decrease with increasing depth. Pressure, temperature and water content of the rocks *in situ* all differ from laboratory conditions. Bullard and Niblett (1951) have discussed these effects and from this it appears unlikely that even their combined influence is sufficient to account for the observed effects. (ii) The borehole might be non-vertical. As far as can be ascertained the borehole lies very close to the vertical throughout its length. (iii) The heat flow in this region might not be vertical. That the direction of the flow lines can be influenced by geological structure and by the angles of the bedding planes has been observed (Bullard 1945) and demonstrated in the laboratory by electrical analogues.

For the lava section between 1305 and 4555 ft, the equation that fits the observed temperatures almost exactly is $T = 20{\cdot}16 + 3{\cdot}177 \times 10^{-3}D + 8{\cdot}02 \times 10^{-8}D^2$. The closeness of the fit is indicated by the difference between the observed temperatures and those calculated from this equation, given in column 5 of table 2(*b*). The deviation of this curve from a straight line is small and would be detected only by temperature measurements to better than 0·1°c. A tangent drawn through the mid-point lies about 0·2°c from the curve at its limits.

Assuming that the lava is thermally isotropic and that the heat flow in this region has been distorted, the variation of vertical heat flow can be found from the slope of the curve at various depths (D is the depth in the borehole):

$$\frac{dT}{dD} = 3{\cdot}177 \times 10^{-3} + 16{\cdot}04 \times 10^{-8}D.$$

Using $K = 0{\cdot}0072$ in equation (1), the vertical heat flow at 1305 ft is

$H_{1305} = 0{\cdot}80 \times 10^{-6}$ cal cm^{-2} sec^{-1}, and at 4555 ft $H_{4555} = 0{\cdot}92 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

The standard error of the mean conductivity is $\sigma(\bar{k}) = 1{\cdot}3 \times 10^{-4}$, and using this to calculate the standard error in the above values of H, taking the temperature gradient as being precisely known, we get $\sigma(H) = 0{\cdot}02$ cal cm^{-2} sec^{-1}.

The geothermic steps in lava near the top and bottom of this borehole indicate (assuming unchanged lava conductivity) a variation in vertical heat flow over a depth of 5500 ft of 15% either side of a mean value of $0{\cdot}86 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

3.3. *Messina Borehole* 0–32

Borehole 0–32 is located in the Northern Transvaal near Messina. There are two features of this area worthy of mention: it is a low-lying part of the country, being 1700 ft above sea level whereas the plateau which covers a large proportion of Southern Africa is about 4000 to 5000 ft; the granitic rocks, which are generally more radioactive than other types, might lead to an increased heat flow in this area. The borehole passes through many narrow bands of granite, dyke and various quartzites, such as quartz–iron–rock, epidotized and quartzose, partly brecciated quartzite and quartzite–epidote–hornblendic rock. Only two sections have a reasonable depth of homogeneous rock, viz. granite from 335 to 489 ft and dyke from 509 to 769 ft.

The diameter of this borehole core is $\frac{7}{8}$ in. and hence too small for the $1\frac{3}{8}$ in. divided-bar conductivity apparatus. Eight rock samples were taken from a nearby

mine and specimens were cut from these in such a manner that the heat flow in the apparatus was in the same direction as *in situ*. The thermal diffusivity of five 3 in. length samples of the borehole core was determined by Mr. Tucker of the Chamber of Mines Research Laboratories and the conductivity was then derived from density and specific heat measurements.

The temperatures were measured one month after drilling to a depth of 1200 ft had been completed, which may not have been sufficient time for temperature equilibrium to be restored. However, even if the temperatures had not quite settled down it is certain that the accuracy of the heat flow calculations is limited by the small number of conductivity measurements. Great care was taken in carrying out the temperature survey as it was desired to see what precision could be attained. All observations were repeated on different days, sometimes in the reverse order to that in which they were made previously. Column 2 of table 2 (*c*) gives the standard error of the means of the observed temperatures.

Heat flow.

(*a*) In the granite from 335 to 489 ft:

geothermic step = 133 ft deg^{-1}

$\bar{k} = 0{\cdot}0057$ (4 samples), $\sigma(\bar{k}) = 5{\cdot}0 \times 10^{-4}$

$H = 1{\cdot}41 \times 10^{-6}$ cal cm^{-2} sec^{-1}, $\sigma(H) = 0{\cdot}12 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

(*b*) In the dyke from 509 to 769 ft:

geothermic step = 122 ft deg^{-1}

$\bar{k} = 0{\cdot}0051$ (2 samples), $\sigma(\bar{k}) = 3{\cdot}5 \times 10^{-4}$

$H = 1{\cdot}37 \times 10^{-6}$ cal cm^{-2} sec^{-1}, $\sigma(H) = 0{\cdot}10 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

Combining all the temperature and conductivity data from the granite and dyke sections between 299 and 797 ft, and making allowance for the small thicknesses of other rocks between these limits, the solution to equation (3) is

$$T = 28{\cdot}22 + 4{\cdot}18 \times 10^{-5} \Sigma(\Delta D/k).$$

The differences between observed and calculated temperatures are very small, as shown in column 4 of table 2 (*c*). This solution gives $H = 1{\cdot}37 \times 10^{-6}$ cal cm^{-2} sec^{-1}, $\sigma(H) = 0{\cdot}09 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

3.4. *Kestell Borehole No. 7*

Borehole No. 7 is near Kestell in the southern Orange Free State. It was drilled to 5000 ft, being completed in 1946. Temperatures were measured by the Geological Survey with an electronic thermometer (Guelke *et al.* 1949) soon after its completion and before its accuracy had been improved. Blockage of the borehole prevented better measurements from being made later. The conductivity of 22 samples of rock core from this borehole was determined by Mossop and Gafner (1951).

Heat flow.

The solution of $T = T_0 + H\Sigma(\Delta D/k)$ and the residuals were calculated by Mossop. The second column of table 2 (*d*) shows the residuals to be relatively large and gives some idea of the order of uncertainty of the readings. The solution is $T = 17{\cdot}17 + 4{\cdot}24 \times 10^{-5}\Sigma(\Delta D/k)$, from which $H = 1{\cdot}39 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

The borehole passes through many narrow bands of sandstone, dolerite and shale, which have a conductivity of 0·0047, 0·0050 and 0·0057 respectively. The temperature–depth curve in figure 2 is reasonably close to a straight line for the whole depth, which is not surprising since the conductivities are about the same. Using the mean conductivity ($\overline{k}$) and mean geothermic step ($\overline{G}$) for the whole depth, $H = 1{\cdot}29 \times 10^{-6}$ cal cm^{-2} sec^{-1}, which indicates that these means are reasonable values to use for estimating the standard error in H. This cannot be done precisely as the temperatures and conductivities are related. Taking $H = \{\overline{k} \pm 2\sigma(\overline{k})\}/\{\overline{G} \pm 2\sigma(\overline{G})\} = (1{\cdot}29 \pm 0{\cdot}22) \times 10^{-6}$ cal cm^{-2} sec^{-1} we have finally $H = 1{\cdot}29 \times 10^{-6}$ cal cm^{-2} sec^{-1}, $\sigma(H) \leqslant 0{\cdot}11 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

§ 4. Heat Flow calculated from Temperature Measurements made by the Geological Survey

The Geological Survey has carried out an extensive temperature survey of Free State and Transvaal boreholes using an electronic thermometer (Bouwer 1953). Use of these temperatures and average rock conductivities has been made to calculate the heat flow.

4.1. *Accuracy of Temperature Measurements*

Laboratory and field checks on the electronic thermometer have indicated that the temperatures are not in error by more than 0·5°c. For example, the electronic thermometer results at 26 points in HB 15 differ by more than 0·2°c from the mercury thermometer results at only four points. The maximum difference is 0·5°c.

An analysis of the electronic thermometer measurements was made to see whether constant errors at given temperatures could be found. The method was to fit the least-squares straight lines to the observed points of temperature against depth for the lava sections of ten boreholes. From the equations to these straight lines the temperature corresponding to various depths was calculated. The differences between calculated and the observed temperature θ were then tabulated in the manner demonstrated in table 3, which is an extract from the complete table. The errors at all temperatures proved to be quite random.

Table 3. Electronic Thermometer Errors: Differences between θ and Calculated Temperatures

Borehole	HB 15	DT 1	TV 2	BU 1	HB 3	VDH 4a
θ=26°c	−0·02		+0·23		+0·01	+0·13
θ=34°c		+0·35	−0·22	−0·07		

4.2. *Heat Flow*

The heat flow values given in table 5 were calculated from the mean of all available rock conductivity data and the slopes of the least-squares straight lines fitted to Bouwer's temperature–depth points for homogeneous sections. The values of thermal conductivity given in table 4 are the mean of all the values given by Mossop and Gafner (1951) and of those given in table 1. Because of the dependence of the conductivity of quartzite on its composition, this mean value will probably not be applicable to all boreholes. Allowance is made for this in the estimated standard errors given as $\sigma'(H) \geqslant$ standard error in H.

Table 4. Mean Rock Conductivities

Rock	Thermal cond.	No. of samples	$\sigma(\bar{k})$
Ventersdorp lava	0·0073	43	2×10^{-4}
Witwatersrand quartzite	0·0136	75	6×10^{-4}
Dolomite	0·0122	15	10×10^{-4}

4.3. *Summary of Heat Flow Results*

All heat flow results given in this paper are to be found in table 5 or table 6. For comparison, other values that have been obtained for the heat conduction to the earth's surface are summarized in table 7. The exceptionally high value found in Nottinghamshire has been attributed to water movement.

Table 5. Heat Flow in Transvaal and Orange Free State (from Bouwer's temperatures)

Borehole			Ventersdorp lava		Witwatersrand quartzite		Dolomite	
			H	$\sigma'(H)$	H	$\sigma'(H)$	H	$\sigma'(H)$
Transvaal (Klerksdorp area)	HB	3	1·15	0·08	1·25	0·11	0·90	0·35
	HB	8	1·12	0·10	—	—	—	—
	HB	14	0·90	0·10	1·19	0·07	0·95	0·23
	HB	15	1·01	0·06	1·22	0·13	—	—
	ST	9	0·98	0·07	1·16	0·17	0·90	0·41
	ST	11	0·91	0·12	—	—	0·30	—
	ST	14	0·84	0·07	—	0·13	0·62	0·41
	R	57	0·86		—	—	—	—
	BU	1	1·08	0·07	1·23	0·13	1·23	0·13
Orange Free State (Northern)	VDH	4a	1·16	0·08	1·25	0·14	—	—
	DT	1	1·05	0·08	1·25	0·12	—	—
	DW	1	1·01	0·13	1·20	0·25	—	—
	TV	2	1·07	0·07	1·27	0·10	—	—
	M	2	1·02	0·11	1·22	0·12	—	—

H is expressed in 10^{-6} cal cm^{-2} sec^{-1}. $\sigma'(H) \geqslant$ standard error in H.

Table 6. Summary of Heat Flow Results

Borehole	Vertical heat flow H	Standard error in H
HB 15	1·05	0·03
Roodepoort	0·80 at 1305 ft	0·02
	0·92 at 4555 ft	0·02
Messina	1·37	0·09
Kestell	1·29	$\leqslant$0·11

H in 10^{-6} cal cm^{-2} sec^{-1}.

Table 7. Other Heat Flow Results

Place	Heat flow (10^{-6} cal cm^{-2} sec^{-1})	Observer
Great Britain	0·98 (±0·12)	Benfield (1939)
California	1·13	Benfield
Michigan	0·93 ($\sigma(H)$=10%)	Birch (1954)
Transvaal	1·16 (±0·09)	Bullard (1939)
Floor of Atlantic Ocean	0·98	Bullard (1954)
England, Nottinghamshire	2·24	Bullard & Niblett (1951)
Yorkshire	1·16	
Persia	0·7 to 1·3	Coster (1947)
Ontario and Quebec	0·69 to 1·05	Misener, Thompson & Uffen (1951)
Floor of Pacific Ocean	1·2	Revelle & Maxwell (1952)

§ 5. Discussion of Results

The heat flow measured in Ventersdorp lava shows a 16% variation either side of $1{\cdot}00 \times 10^{-6}$ cal cm^{-2} sec^{-1}. The variation is in all probability real and to be accounted for in terms of geological structure. There is definite evidence that in the Roodepoort borehole (R 57) the vertical heat flow differs considerably from top to bottom. ST 11 and ST 14, which are both near to R 57, show correspondingly low values of H in the lava, which is at about the same depth.

H values measured in Witwatersrand quartzite show remarkably little spread considering its heterogencity. On the average the quartzite is deeper than the lava and it may be that in consequence the heat flow is less influenced by the topography and varied structure near the surface of the earth. The heat flow values are consistently higher than in lava; this is most likely due to the doubtful value of the mean conductivity used. If, for example, the figure of 0·0128 given in table 1 had been used instead of the mean of all available results, the agreement would have been improved. There is, however, no justification for doing this.

The results in dolomite serve to emphasize that indiscriminate measurements of H can lead to large errors. The fissured and sometimes even cavernous nature of dolomite permits movement of water and hence the disturbance of temperatures. Measured conductivities spread over a wide range. In some cases the dolomite layer is thin and in all cases given here it overlies the quartzite and lava.

In conclusion it can be stated that existing measurements show that the average heat flow varies little from place to place in South Africa. Measurements at points scattered over the earth's surface mostly lie w thin 40% of the figure $1{\cdot}0 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

Acknowledgments

This paper is published with the permission of the President of the South African Council for Scientific and Industrial Research.

I wish to acknowledge with thanks the assistance of the following: Mr. S. C. Mossop who initiated a greal deal of the work, Messrs. G. Gafner and G. J. Viljoen who carried out the rock conductivity measurements and many of the calculations, the Geological Survey Office, Messina (Transvaal) Development Co. Ltd., and the Anglo-Transvaal Consolidated Investment Co. Ltd. for furnishing borehole logs and rock cores and assistance in connection with the field work, Mr. Tucker of the Chamber of Mines Research Laboratories for doing diffusivity measurements and Dr. H. S. Steyn for advice on statistical matters.

References

Benfield, A. E., 1939, *Proc. Roy. Soc.* A, **173,** 428.
Birch, F., 1954, *Amer. J. Sci.*, **252,** 1.
Bouwer, R. F., 1953, *Trans. Geol. Soc. S. Afr.*, in the press.
Bullard, E. C., 1939, *Proc. Roy. Soc.* A, **173,** 474; 1945, *Nature, Lond.*, **156,** 35; 1954, *Proc. Roy. Soc.* A, **222,** 408.
Bullard, E. C., and Niblett, E. R., 1951, *Mon. Not. R. Astr. Soc. Geophys. Suppl.*, **6,** 222.
Coster, H. P., 1947, *Mon. Not. R. Astr. Soc. Geophys. Suppl.*, **5,** 131.
Guelke, R., Heydenrych, J. C. R., and Anderson, F., 1949, *J. Sci. Instrum.*, **26,** 150.
Krige, L. T., 1948, *Geological Series Bulletin No.* 18, Dept. of Mines, Union of S. Africa.
Misener, A. D., Thompson, L. G. D., and Uffen, R. S., 1951, *Trans. Amer. Geophys. Un.*, **32,** 729.
Mossop, S. C., 1950, *J. Chem. Soc. S. Afr.*, **51,** 120.
Mossop, S. C., and Gafner, G., 1951, *J. Chem. Soc. S. Afr.*, **52,** 61.
Revelle, R., and Maxwell, E. A., 1952, *Nature, Lond.*, **170,** 199.

The Order of Coincidence Required for Visual Threshold

By G. S. BRINDLEY
R.A.F. Institute of Aviation Medicine, Farnborough, Hants

MS. received 15*th June* 1954

Abstract. It is proved that the limiting hypothesis of an infinite number of identical independent units each depending for excitation on the scoring of a twofold quantum coincidence implies a steeper frequency-of-seeing curve than any other strictly two-quantum hypothesis and than the quasi-two-quantum hypothesis suggested by Bouman and van der Velden to account for large-field frequency-of-seeing curves. Published experimental curves for large and long flashes are steeper than those implied by the limiting two-quantum hypothesis or by the corresponding three-quantum hypothesis. It follows that the order of coincidence upon which the threshold for such flashes depends is at least fourfold.

Hecht, Shlaer and Pirenne (1942) and van der Velden (1944) have inferred from the frequency-of-seeing curve for a short flash covering a small area of retina the number of quanta which must be absorbed in receptors for the flash to be seen, assuming that the absorption of a fixed small number of quanta will cause a light sensation whatever their position and instant in the small brief flash. If I (a multiple of the light intensity) is the mean number of quanta absorbed, and P the probability that m or more will be absorbed, then

$$P=1-e^{-I}(1+I+I^2/2!+I^3/3!+\ldots \text{ to } m \text{ terms}).$$

The inferences made in the two early papers from small-field frequency-of-seeing curves disagree, mainly because of differences between the experimental results obtained, those of van der Velden being consistent with $m=2$ and those of Hecht, Schlaer and Pirenne with $m=5$ or more. The sources of the disagreement have been discussed by Pirenne and Denton (1951).

For large and long flashes the experimental results of Bouman and van der Velden (1947) and of Denton and Pirenne (1954) agree well, and a clear inference can be made from them.

Let us suppose first that for a large and long flash there are many, say N, identical independent units (defined in time as well as in space) such that, if at least one of them is stimulated by it, the flash will be seen. Let the probability of excitation of a given unit be P and the probability of seeing be Q. Then

$$Q=1-(1-P)^N=1-e^{-IN}(1+I)^N$$

if excitation of each unit depends on a twofold quantum coincidence,

or

$$1-e^{-IN}(1+I+I^2/2)^N$$

if it depends on a threefold coincidence.

As $N\to\infty$ these relations tend in the limit to

$$Q_0=1-\exp(-NI^2/2) \text{ for a twofold coincidence,}$$

or

$$Q_0=1-\exp(-NI^3/6) \text{ for a threefold coincidence.}$$

For these limiting values $dQ/d\log I$ for a given value of Q is greater than for any other hypothesis involving the same order of coincidence. In particular, taking $m=2$, we have:

(1) If N is finite, then $Q_1=1-e^{-IN}(1+I)^N$.

(2) If, owing to biological variations, there is a probability $\alpha\neq 1$ of seeing when the two-quantum condition is met and a probability $\beta\neq 0$ of seeing when it is not met, then $Q_2=\alpha Q_0+\beta(1-Q_0)$.

(3) If, as Bouman and van der Velden (1947) suggest, the falling of a third quantum within a specified space–time region γ times the size of the independent unit decreases the probability of seeing when two others fall within the unit from 1 to a value δ, where $1>\delta\geqslant 0$, then

$$Q_3=[1-e^{-IN}(1+I)^N][e^{-\gamma I}+\delta(1-e^{-\gamma I})].$$

(4) If *either* (*a*) the independent units are not all identical, but differ from each other in summation time, summation area and sensitivity per unit area and time, *or* (*b*) the units are not wholly independent, so that the probability of seeing, which is unity if two or more quanta are absorbed in *the same* unit has some value ϵ $(1>\epsilon>0)$ if two quanta are absorbed in *different* units, where ϵ for a given pair of units may be a function of their spatial or temporal separation, but is independent of the absorption of quanta in other units, then

$$Q_4=1-\prod_r[\exp(-k_rIN_r)(1+k_rI)^{N_r}]$$

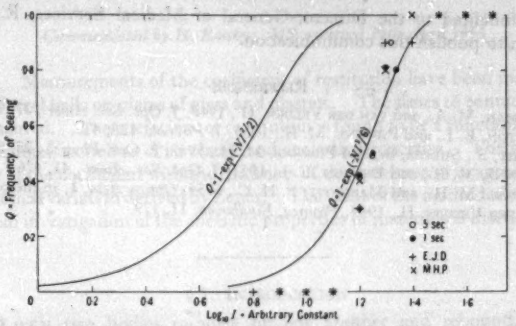
where N_r is the number of units in the class r, and k_r the sensitivity of a member of this class.

Since in hypothesis 4(*b*) each pair of units constitutes an independent subsidiary unit, 4(*b*) is a special case of 4(*a*).

For all these hypotheses it can be proved that $dQ_0/d\log I>dQ_s/d\log I$ at the value $Q_0=Q_s(s=1,2,3$ or $4)$. The proofs for Q_4 and Q_3 are given in the Appendix. That for Q_2 is almost trivial, and Q_1 is a special case of Q_4. It follows that any experimental frequency-of-seeing curve whose gradient at each value of Q exceeds that of $Q=1-\exp(-NI^2/2)$ excludes absolutely a twofold quantum coincidence as determining threshold under the conditions of the experiment. The corresponding inequalities can be proved for $m=3$, so that an experimental curve steeper than $Q=1-\exp(-NI^3/6)$ excludes a three-quantum coincidence.

The frequency-of-seeing values obtained by Bouman and van der Velden (1947) and by Denton and Pirenne (1954) for large and long flashes are plotted in the figure, together with the limiting theoretical curves for twofold and threefold coincidences. It will be seen that the experimental gradient exceeds that of either theoretical curve. Hence the order of quantum coincidence concerned in reaching threshold for a large and long flash is at least fourfold. The units registering these coincidences must be responsible for threshold also for any flash larger and longer than their summation time and area; and flashes smaller and shorter than this can have their threshold determined by twofold coincidences only if the units registering these have very much smaller summation regions, so that, although for small flashes they are more sensitive than the units registering higher orders of coincidence, for large flashes they are insufficiently sensitive to contribute even to the low-probability tail of the frequency-of-seeing curve.

Though it is not at all unlikely that units differ greatly in the size of their summation region, there is little positive evidence to supprt a hypothesis that the small units score lower orders of coincidence than the large, since the most precise of the arguments of van der Velden for twofold coincidences depends on spatial considerations which are largely invalidated by the spatial variations in sensitivity found by Pirenne and Marriott (1954), as well as by the present arguments.



Limiting theoretical large-field frequency-of-seeing curves for $m=2$ and $m=3$, with the experimental results of Bouman and van der Velden (circles) for a 23° 10′ field and 5-second and 1-second exposures, and of Denton and Pirenne for a 45° field and a 5-second exposure (crosses).

APPENDIX

$$\frac{dQ_0}{d\log I} = -2(1-Q_0)\log(1-Q_0)$$

$$\log(1-Q_4) = \sum_r [-k_r I N_r + N_r \log(1+k_r I)]$$

$$= -\sum_r [N_r(\tfrac{1}{2}k_r{}^2 I^2 - \tfrac{1}{3}k_r{}^3 I^3 + \tfrac{1}{4}k_r{}^4 I^4 - \ldots)],$$

$$\frac{dQ_4}{d\log I} = -(1-Q_4)\frac{d}{d\log I}[\log(1-Q_4)]$$

$$= -(1-Q_4)\sum_r \left[-k_r N_r + \frac{k_r N_r}{1+k_r I}\right] I$$

$$= -(1-Q_4)\sum_r [N_r(-k_r{}^2 I^2 + k_r{}^3 I^3 - k_r{}^4 I^4 + \ldots)]$$

$$= -(1-Q_4)\{2\log(1-Q_4) + \sum_r [N_r(\tfrac{1}{3}k_r{}^3 I^3 - \tfrac{3}{4}k_r{}^4 I^4 + \tfrac{3}{5}k_r{}^5 I^5 - \ldots)]\}$$

$$< dQ_0/d\log I \quad \text{when} \quad Q_4 = Q_0$$

$$\frac{dQ_3}{d\log I} = [1-e^{-IN}(1+I)^N][-\gamma(1-\delta)e^{-\gamma I} I]$$

$$+ [e^{-\gamma I} + \delta(1-e^{-\gamma I})]\frac{d}{d\log I}[1-e^{-IN}(1+I)^N]$$

$$< dQ_0/d\log I \quad \text{when} \quad Q_3 = Q_0 \quad \text{since}$$

$$\frac{d}{d\log I}[1-e^{-IN}(1+I)^N] < \frac{dQ_0}{d\log I},$$

for this is the special case Q_1 of the inequality proved above, and

$$1>[e^{-\gamma I}+\delta(1-e^{-\gamma I})]>0$$
$$1>[1-e^{-IN}(1+I)^N] >0$$
$$0>[-\gamma I(1-\delta)e^{-\gamma I}] >-1.$$

Acknowledgments

I am indebted to the Director-General of Medical Services, R.A.F., for permission to publish this communication.

References

Bouman, M. A., and van der Velden, H., 1947, *J. Opt. Soc. Amer.*, **37**, 908.
Denton, E. J., and Pirenne, M. H., 1954, *J. Physiol.*, **123**, 417.
Hecht, S., Shlaer, S., and Pirenne, M. H., 1942, *J. Gen. Physiol.*, **25**, 819.
Pirenne, M. H., and Denton, E. J., 1951, *J. Opt. Soc. Amer.*, **41**, 426.
Pirenne, M. H., and Marriott, F. H. C., 1954, *Optica Acta*, **1**, in the press.
van der Velden, H., 1944, *Physica, Eindhoven*, **11**, 179.

A Study of the Impact on Spheres of Plates

By J. P. A. TILLETT
Imperial Chemical Industries Limited, Butterwick Research Laboratories, Welwyn, Herts.

Communicated by H. Kolsky; *MS. received 29th April* 1954

Abstract. Measurements of the coefficient of restitution have been made for the impact of steel balls on plates of glass and plastics. The times of contact have also been measured. The coefficient of restitution is found to be a function of the size of the specimen as well as of its elastic and anelastic properties. The variation in the value of the coefficient with the thickness of the specimen is in agreement with the theoretical variation derived by Zener. The value of the method of rebounding balls for an investigation of the anelastic properties of materials is discussed.

§ 1. Introduction

When two bodies impinge on one another and rebound, the total momentum must be conserved, but some of the kinetic energy of the bodies will be changed to other forms of energy, and ultimately to heat, in two principal ways. First, elastic vibrations will generally be set up so that some energy goes into internal motion, and secondly, since the colliding bodies will not be perfectly elastic, energy will be lost in viscous and plastic effects in the region of contact.

The motion of bodies after impact depends on the forces acting in the region of contact, and the use of an assumption introduced by Hertz (1881) makes it possible to obtain an approximate solution to problems of elastic impact which it would be difficult to solve rigorously. The assumption is that, in the region near the point of contact, the stresses and strains may be computed, at any instant, as though the contact were static. This is valid if the distance travelled by the elastic waves during impact is very large compared with the dimension of the area of contact. The effect of the elastic waves within the bodies was first mentioned by Saint-Venant (1867) for the impact of long rods, and he showed that, as a consequence, even perfectly elastic materials will have a coefficient of restitution less than unity.

A successful application of the Hertz theory has been to the impact of rods with rounded ends. The kinetic energy loss due to elastic waves within the bars has also been taken into account and the agreement with experiment is satisfactory (Sears 1908, Wagstaff 1924). The validity of Hertz's assumption has been confirmed in other particular cases by Andrews (1931) and by Crook (1952).

Another case in which the forces acting in the region of contact and the vibrations within the material may be considered together is for the impact of spheres on plates. Experiments of this type were made by Raman (1920), who also gave a discussion of the results. A more elaborate treatment of the theoretical problem involved was given by Zener (1941). Using the premise of Hertz, he found the amount of energy transferred from the ball to flexural vibrations of the plate.

Although for a complete treatment it would be necessary to consider all the modes of vibration, for a thin plate the most important are those of flexure, so that Zener's theory gives a good approximation.

The method of rebounding balls has been used in various ways for investigating the dynamic elastic properties of materials. Work on metals has been carried out by Davies (1949), who has measured static and dynamic yield strengths using steel balls, and by Tabor (1948), who has related dynamic hardness to elastic constants. Jenckel and Klein (1952) have used rebounding balls to determine relaxation-time spectra of polymers.

The experiments described below are similar to those of Raman. Steel balls were bounced on rectangular plates of different materials and the coefficient of restitution was measured, varying the size of the ball, dimensions of the plate and other parameters. The time of contact was also measured. The purpose of the work was to investigate the condition under which Hertz's assumption is valid and the ideas of Saint-Venant apply and to confirm Zener's theory. Discrimination can then be made between effects due to the size and shape of the specimen and those due to the anelastic properties of the materials.

§ 2. Theoretical

The coefficient of restitution is defined here as the ratio of the velocities of the sphere, after and before impact. If h_1 is the height to which the ball rebounds when dropped from a height h_0, the coefficient of restitution e for the particular ball and specimen is given by $e=(h_1/h_0)^{1/2}$. If the coefficient of restitution were independent of the velocity of impact, a plot of the logarithm of the height of successive bounces against the number of bounces that have occurred would be a straight line.

The analysis of the reaction of large plates to impulsive forces acting at a point has been applied, by Zener, to the impact of spheres on plates. The relation between the displacement of the point of application in the plate U and the impulse P is of the form $U=\alpha P$, where α is obtained from the solution, given by Zener, of the differential equation for the transverse displacement of the plate under the action of an impulsive force. In the application of this theory, the equations of motion of the sphere and the plate are solved simultaneosuly. The relation between the reaction F of the plate on the sphere and their relative displacement s is $F(s)=ks^{3/2}$, where k is a constant involving the radius of the sphere and the elastic constants of the plate and sphere. This is the expression established by Hertz for bodies in contact and used in deriving an expression for the time of contact and pressure–time relation for perfectly elastic bodies.

The relation between relative displacement and time obtained by Zener is

$$\frac{d^2\sigma}{d\tau^2}+\left(1+\lambda\frac{d}{d\tau}\right)\sigma^{3/2}=0 \qquad \ldots\ldots(1)$$

where σ, τ are dimensionless variables: σ is proportional to the relative displacement of the plate and sphere and τ to the time from the beginning of impact; λ is a dimensionless quantity which depends on the parameters of impact and has been termed by Zener the *inelasticity parameter*. The value of λ is given by

$$\lambda=\frac{\pi^{3/5}}{3^{1/2}}\left(\frac{R}{2h}\right)^2\left(\frac{v_0}{v'}\right)^{1/5}\left(\frac{\rho_1}{\rho_2}\right)^{2/5}\left(\frac{E_1'}{E_1'+E_2'}\right)^{2/5}$$

where R is the radius of the ball, $2h$ the thickness of the plate, v_0 the velocity of impact and ρ_1, ρ_2 the densities of the ball and plate. If ν_1, E_1 and ν_2, E_2 are Poisson's ratio and Young's modulus of the materials of the ball and plate, then

$$E_1' = \frac{E_1}{1-\nu_1^2}, \qquad E_2' = \frac{E_2}{1-\nu_2^2} \qquad \text{and} \qquad v' = (E_2'/\rho_2)^{1/2}.$$

When σ returns to zero the coefficient of restitution has the value of $d\sigma/d\tau$ and the time of contact is proportional to the value of τ. The coefficient of restitution may be evaluated by a numerical integration of equation (1) for different values of λ.

For $\lambda = 0$, $e = 1$ and the time of contact and pressure–time relation are as given by Hertz. The time of contact in this case is (Love 1927, p. 199)

$$T_H = 2{\cdot}94\{(5/4)\pi\rho_1(1/E_1' + 1/E_2')\}^{2/5} R/v_0^{1/5}. \qquad \ldots\ldots (2)$$

Many materials cannot be regarded, even approximately, as perfectly elastic, and there will be energy losses due to the relaxation of the stress during impact. The deformation of the plate in the region of contact and the recovery as the ball rebounds may be considered as half a cycle of a periodic disturbance. The case considered is that of a plate so thick that flexural vibrations are negligible. If the original energy of the ball is E, equal to $\frac{1}{2}mv_0^2$, and the kinetic energy loss due to anelastic effects is ΔE,

$$\Delta E = \tfrac{1}{2}mv_0^2(1-e^2), \qquad \Delta E/E = (1-e^2).$$

If Δ' is the logarithmic decrement defined as the logarithm of the ratio of the amplitudes of a disturbance on opposite sides of a mean position and Δ the corresponding value for consecutive amplitudes on the same side of a mean position, then

$$\Delta' = \tfrac{1}{2}\Delta E/E \qquad \text{and} \qquad \Delta = 2\Delta' = (1-e^2).$$

In this instance the logarithmic decrement that is measured is for a deformation which is due to both shear and compression. It will be very nearly equal to the value for pure shear if the energy stored in the form of compression is small compared with that stored in shear. A calculation shows that for the Hertz distribution, where the relevant parameter involving the elastic constants of the material is $E/(1-\nu^2)$, the ratio of the energies stored in shear and compression is

$$\frac{2(1-\nu+\nu^2)}{(1-2\nu)(1+\nu)}.$$

If Poisson's ratio has a value 0·35, 80% of the energy is stored in the form of shear, and if $\nu = 0{\cdot}25$ the amount is 72%.

In order to compare results with those obtained by other methods it is desirable to consider the loss which occurs during impact as a loss at a definite frequency. In similar work on polymers, Jenckel and Klein have regarded the reciprocal of twice the time of contact as a suitable frequency but did not give a detailed derivation of their approximation. In the Appendix to this paper the connection between the energy loss corresponding to an 'equivalent frequency' is discussed. It is shown that the energy loss during impact may be attributed to loss at an 'equivalent frequency' which has a value very nearly equal to that assumed by Jenckel and Klein. With most materials, however, the variation of logarithmic decrement with frequency is not rapid, so that it is not necessary to know the value of the frequency exactly.

In considering the kinetic energy loss of the ball during impact, due to the elastic properties of the materials, the energy in the elastic wave motion has been

ignored. With glass and hard steel specimens the total energy loss of a ball on rebound is of the order of 1%. The energy in the elastic waves in these materials must necessarily be small and, since this energy depends only on the elastic constants, it is not expected that it will be of a very different order of magnitude in other materials.

§ 3. Experimental

3.1. *Measurement of the Coefficient of Restitution*

The coefficient of restitution of a ball bouncing on a plate of the material under examination was obtained by measurement of the height to which the ball rebounded when dropped from a known height. The motion of the ball was recorded photographically and, by allowing the ball to bounce a number of times in one experiment, the coefficient of restitution at different velocities of impact was obtained.

The plate rested on a levelling table which was adjusted at an angle so that the ball bounced at a constant distance from the camera and across its field of view. In this way the traces of successive bounces were separated from one another on the photographic plate and could be easily measured using a Beck travelling microscope with traverse in two directions at right angles.

The coefficient of restitution for glass and 'Perspex' (a plasticized form of polymethyl methacrylate) samples of different thicknesses was measured, using steel balls ranging in size from $\frac{1}{16}$ in. to $\frac{1}{2}$ in. in diameter. The glass plates varied in thickness from 0·036 in. to 0·986 in. and the Perspex plates from 0·15 in. to 4·45 in. Measurements were also made for Perspex plates of different thicknesses when backed by a massive steel cylinder with grease between the surfaces of the anvil and the sample. For steel, the coefficient of restitution was measured at lower velocities of impact by suspending a ball by a fine thread and swinging it horizontally against the specimen. The velocity of impact was such as to avoid the production of any permanent indentation.

3.2. *Variation of Coefficient of Restitution with Temperature*

The way in which the coefficient of restitution varies with temperature for the impact of a steel ball on thick plates of Perspex and on a block of an organic glass, glycerol sextol phthalate (a mixed glycerol-cyclohexanol ester of phthalic acid) has been investigated. The specimen was mounted in a thermostatically controlled vessel which had a Perspex window in one side so that the motion of the ball could be photographed. Light from a small source (a 'Pointolite') was reflected on to the ball by an inclined mirror within the enclosure. The temperature was varied between 18 and 90°C for Perspex and between 18 and 60°C for glycerol sextol phthalate.

3.3. *Measurement of the Time of Contact*

For glass and Perspex samples the variation of the time of contact with changes in velocity of impact, size of ball and thickness of specimen was investigated. A thin layer of colloidal graphite was sprayed on to the surface of the specimen and a connection made to earth potential. A potential of 10 volts was applied to a ball, from a battery, by means of a fine wire soldered to it which also acted as a suspension. The potential difference between the surface of the specimen and the ball was amplified and fed on to one pair of Y deflector plates of a high-speed, double-beam

cathode-ray oscillograph. When contact was made between the ball and the surface the potential difference was reduced and the change was visible on the oscillograph trace. This value was maintained until the ball left the surface. A block diagram of the circuit is shown in figure 1.

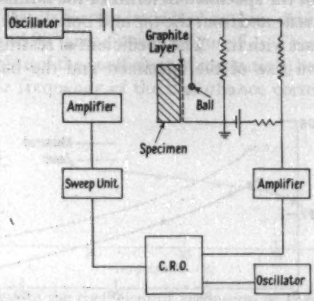


Figure 1. Block diagram of the circuit used for measuring the time of contact.

It was necessary that the time sweep of the oscillograph should be triggered just before contact was made. The ball was therefore connected to the grid of an LC valve oscillator so that the change in capacity between the ball and plate as the surface was approached brought an off-tuned circuit to resonance. A triggering pulse was provided from the anode circuit and this was also used to initiate a brightening pulse for the cathode-ray oscillograph. A 100 kc/s oscillator was connected to the second pair of Y plates to provide a timing calibration.

§ 4. Results

(*a*) The coefficient of restitution was measured at 20°c with the specimen sufficiently thick for its size not to affect the value obtained. If the time of contact is greater than the time required for elastic waves to return to the point of impact, after having been reflected at the bottom face, then the thickness of the specimen is effectively infinite.

Measured values of the coefficient of restitution are: glass 0·985, Perspex 0·950, glycerol sextol phthalate 0·940, steel 0·950.

(*b*) The variation of the coefficient of restitution with the velocity of impact may be found, in any one experiment, from a plot of the logarithm of the height of rebound against the number of bounces that have taken place. When the plates are thick, relative to the diameter of the ball, there is only a small variation with velocity of impact, which is of the order of $\frac{1}{2}\%$ over a velocity range of 300 cm sec^{-1}. With thin plates the variation is much greater. For example, it is 12% for the impact of an $\frac{1}{8}$ in. diameter ball on a glass plate of thickness 0·08 inch. In all cases the rate of change for each successive bounce is decreasing although the rate of change with velocity is greater at low velocities of impact. The magnitude of the variation with velocity that was found experimentally agrees with the value predicted by the theory of Zener (figure 2). An estimate was made of the loss due to air damping and its variation with height of fall, and this was found to be small compared with the variation with velocity already mentioned.

In the results discussed below, the coefficient of restitution corresponds to a velocity of impact of approximately 90 cm sec^{-1}.

(*c*) The effect of the finite thickness of the plate on the value of the coefficient of restitution is shown for Perspex and glass in figures 3 and 4. It is convenient to express the thickness of the specimen in terms of the number of times that a longitudinal wave could be reflected from the top and bottom of the plate during the time that the ball is in contact with it. The coefficient of restitution is reduced by half, because of the relative size of the specimen and the ball, when the thickness

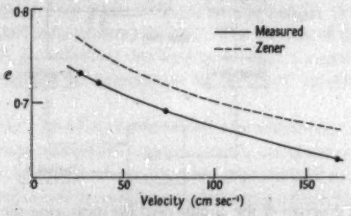


Figure 2. The variation of the coefficient of restitution of a thin glass plate with the velocity of impact.

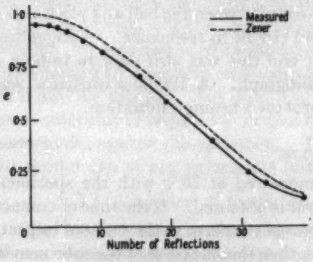


Figure 3. The variation of the coefficient of restitution of Perspex plates with the thickness of the specimen.

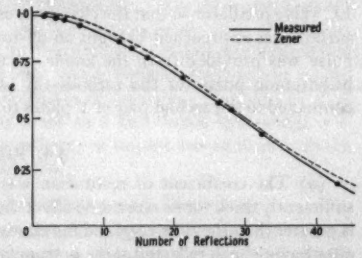


Figure 4. The variation of the coefficient of restitution of glass plates with the thickness of the specimen.

corresponds to about 30 reflections in glass and 20 in Perspex. The fact that the coefficient of restitution is not unity when the plate is so thick that there is not time for even one reflection while the ball is in contact is due to the relaxation of the stress in the material and also to the fact that a small amount of energy goes into the elastic waves set up in the plate. The values for the coefficient of restitution calculated from Zener's formulae are always greater than those obtained experimentally. They differ, however, by an almost constant amount which corresponds to the energy loss not accounted for by Zener. Values of e below 0·1 are not very reliable because of the difficulty in measuring accurately such small heights of rebound.

(*d*) When the Perspex plates are backed by a large steel cylinder the height of rebound does not decrease with thickness in the same way. The coefficient of

restitution appears to have a constant value of 0·955 down to such a thickness that 15 reflections of the longitudinal wave could take place, and falls to 0·855 for a thickness corresponding to 50 reflections.

(*e*) It is proposed to discuss the results obtained for the variation of the coefficient of restitution with temperature in another paper, but the trend will be indicated here. The results for the impact of an $\frac{1}{8}$ in. diameter ball on 2 in. thick Perspex between 20° and 90°C are shown in figure 5. The logarithmic decrement (calculated from the value of the coefficient of restitution) has a value of 0·1 at room temperature, when the frequency of the disturbance corresponds approximately to 13·0 kc/s.

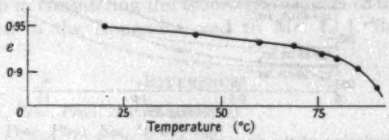


Figure 5. The variation of the coefficient of restitution of Perspex with temperature.

The variation with temperature has been found for glyercol sextol phthalate with several different sized balls in the temperature range 18 to 68°C. It was found that even at 60°C, when the static viscosity was 3×10^5 poises, an $\frac{1}{8}$ in. ball had a coefficient of restitution of 0·5. The ball, however, is only in contact with the material for about 5×10^{-5} second, so that the rate of straining is high. The logarithmic decrement corresponding to the measured value of the coefficient of restitution has been calculated for a $\frac{1}{4}$ in. diameter ball at different temperatures. The value of the 'equivalent frequency' has been discussed above. These results are shown below, together with those obtained by Benbow (1954) for the logarithmic decrement of torsional oscillations.

Logarithmic Decrement from Torsion and Impact Measurements

Temperature (°C)	Torsion (f=1·0 kc/s)	Impact (f=4·4 kc/s)
22	0·10	0·11
26	0·16	0·15
30	0·23	0·22

It might be expected that at higher temperatures the stickiness of the surface would affect results. However, when a very thin sheet of 'Cellophane' was placed on the surface there was no appreciable difference in the value of the coefficient of restitution.

(*f*) The variation of the time of contact with the velocity of impact v_0 in the case of thick plates of glass and Perspex is found to be in agreement with the prediction of Hertz, in that the time of contact is proportional to $1/v_0^{1/5}$. The measured times of contact for the impact of a $\frac{1}{4}$ in. diameter steel ball on plates of different materials with a velocity of 90 cm sec^{-1} are: Perspex 73×10^{-6} sec, glass 32×10^{-6} sec, steel 21×10^{-6} sec. The time of contact with Perspex indicates a value of $5{\cdot}4 \times 10^{10}$ dyn cm^{-2} for Young's modulus at these rates of loading (see equation (2)).

The way in which the time of contact varies with the thickness of the plate for different velocities of impact is shown in figure 6, together with the values deduced from Zener's equation. This gives another indication of the thickness of specimen for which the Hertz assumption is valid. The measured times of contact, however, appear to be greater for thin plates than the values predicted by Zener.

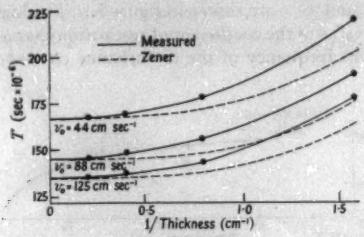


Figure 6. The variation of the time of contact of a $\frac{1}{4}$ in. diameter ball on Perspex with the thickness of the specimen.

§ 5. Discussion

Schnurmann (1941) has considered the impact of steel balls on tool steel and ball-bearing steel anvils and obtained values for the coefficient of restitution at different velocities of impact and with different sized anvils by measuring the total bouncing times. Since the coefficient of restitution varies with the velocity of impact this is an approximate method, but is more accurate the greater the value of the coefficient of restitution. This is the case for impact at low velocities on thick anvils, and under these conditions it was found that the coefficient of restitution was constant and always had a value significantly less than unity. These results, which depend on the relative sizes of the ball and anvil, were observed by Schnurmann and, together with the results described above, are in agreement with the deduction from Zener's theory if allowance is made for the fact that the theory does not attempt to account for the total energy loss of a rebounding ball.

In Raman's experiments on thin plates which were described above, the energy losses, apart from those due to vibrations of the plate, were not apparent because of their small value in glass. From the results shown above for Perspex the magnitude of the kinetic energy loss that would occur for impact on a very large block may be seen. By differentiating between the several mechanisms which cause a loss of the kinetic energy of the ball, this method may be used to obtain a value for the loss due to relaxation of the stress in the region of contact under different conditions. The complications due to the size of the specimen may apparently be avoided by making the plate sufficiently thick or by mounting it in close contact with a massive anvil. By using balls of different sizes it should be possible to obtain values for the logarithmic decrement of the disturbance over a range of frequencies. It is proposed to discuss this in detail later and to describe a method of obtaining a relaxation-time spectrum from such results.

The method of ball impact for investigating the dynamic properties of materials has several inherent advantages. The rate of loading is very high and large stresses are produced near the surface of contact. Measurements may easily be made at

different temperatures. A disadvantage is that the stress cycle involves different types of deformation which have dissimilar frequency spectra. The rate of deformation depends on $v_0^{-1/5}$ and the radius of the ball, so that the variation of frequency that is possible by changing only these parameters is in practice limited. The method would appear to be most useful for materials with moderate damping and for those with a large variation in damping with temperature.

ACKNOWLEDGMENTS

The author wishes to thank Dr. H. Kolsky for his advice and guidance throughout the course of this work. She also wishes to express her thanks to Dr. F. C. Roesler for his help in considering the theoretical aspects of the problem and for the work contained in the Appendix, and to Mr. J. J. Benbow for helpful discussions.

REFERENCES

ANDREWS, J. P., 1931, *Proc. Phys. Soc.*, **43,** 8.
BENBOW, J. J., 1954, *Proc. Phys. Soc.* B, **67,** 120.
CROOK, A. W., 1952, *Proc. Roy. Soc.* A, **212,** 377.
DAVIES, R. M., 1949, *Proc. Roy. Soc.* A, **197,** 416.
HERTZ, H., 1881, *J. reine angew. Math.*, **92,** 156 (see Love, 1927, p. 198).
JENCKEL, E., and KLEIN, E., 1952, *Z. Naturf.*, **7a,** 619.
KOLSKY, H., 1953, *Stress Waves in Solids* (Oxford : Clarendon Press).
LOVE, A. E. H., 1927, *The Mathematical Theory of Elasticity*, 4th Edn (University Press : Cambridge).
RAMAN, C. V., 1920, *Phys. Rev.*, **15,** 277.
ROESLER, F. C., and PEARSON, J. R. A., 1954, *Proc. Phys. Soc.* B **67,** 338.
DE SAINT-VENANT, B., 1867, *J. de Math. (Liouville)*, Ser 2, 12 (see Love, 1927, p. 431).
SCHNURMANN, R., 1941, *Proc. Phys. Soc.*, **53,** 662.
SEARS, J. E., 1908, *Trans. Camb. Phil. Soc.*, **21,** 49.
TABOR, D., 1948, *Proc. Roy. Soc.* A, **192,** 247.
WAGSTAFF, J. E., 1924, *Proc. Roy. Soc.* A, **105,** 544.
ZENER, C., 1941, *Phys. Rev.*, **59,** 669.

APPENDIX

The purpose of this Appendix is to establish a relation between the energy loss due to internal friction during impact and the relaxation spectrum of the material. The relation is based on the assumption that the value of the coefficient of restitution is not very different from unity, so that the Hertz theory may be used to derive the indentation-time function.

The rate of work done by an external periodic force is equal to the product of the components of velocity and force in phase at the point of application. The energy loss is obtained by integrating over a whole number of cycles. If the material is first regarded as behaving as a single Maxwell element (Kolsky 1953, chap. V) with associated moduli E and η, and is subjected to a periodic displacement $x = Ae^{i\omega t}$, then the energy loss in one cycle is given by

$$\overline{W} = -A^2\pi\left\{\frac{\eta\omega}{1+(\eta\omega/E)^2}\right\}. \qquad \ldots\ldots(A\,1)$$

For the case of a real material which has a spectrum of relaxation times, if the deformation is of the form $x=Ae^{i\omega t}$, the force is given by $P=EAe^{i\omega t}$ where $E=E'+iE''$ and

$$E_\omega'=\int_0^\infty \frac{\omega^2T^2}{1+\omega^2T^2}\phi(T)\,dT, \qquad \text{.....(A 2)}$$

$$E_\omega''=\int_0^\infty \frac{\omega T}{1+\omega^2T^2}\phi(T)\,dT. \qquad \text{.....(A 3)}$$

$\phi(T)$ is a function defining the relaxation-time spectrum (cf. Roesler and Pearson 1954). The energy loss in one cycle is then

$$\overline{W}=A^2\pi\int_0^\infty \phi(T)\frac{\omega T}{1+\omega^2T^2}\,dT. \qquad \text{.....(A 4)}$$

The indentation-time function from the Hertz theory of impact may be expressed as a Fourier integral of the form

$$x(t)=\int_{-\infty}^{+\infty} A_\omega e^{i\omega t}\,d\omega$$

where A_ω is the amplitude associated with a pulsatance between ω and $\omega+d\omega$. To consider the energy loss during impact this deformation must be applied to a material with a relaxation-time spectrum and the product of the components of velocity and force in phase integrated over a suitable time interval. The real part of the velocity is

$$\frac{dx(t)}{dt}=-\int_{-\infty}^{+\infty} A_\omega\,\omega\sin\omega t\,d\omega. \qquad \text{.....(A 5)}$$

From the superposition principle the force may be evaluated by summing over all components.

$$P(t)=\int_{-\infty}^{+\infty} x(\omega,t)E_\omega\,d\omega$$

and the real part of $P(t)$ is

$$\int_{-\infty}^{+\infty} d\omega\left\{A_\omega\cos\omega t\int_0^\infty dT\left[\frac{\omega^2T^2}{1+\omega^2T^2}\phi(T)\right]\right\} -\int_{-\infty}^{+\infty} d\omega\left\{A_\omega\sin\omega t\int_0^\infty dT\left[\frac{\omega T}{1+\omega^2T^2}\phi(T)\right]\right\}. \qquad \text{.....(A 6)}$$

$dx(t)/dt$ in equation (A 5) is an odd function of t and the first term of (A 6) is even in t, so that if the range of integration is centred on zero the cross-product of (A 5) and the first term of (A 6) will always be zero. The total work done may be written

$$\overline{W}=\int_{-\infty}^{+\infty} dt\left\{\int_{-\infty}^{+\infty} d\omega A_\omega\,\omega\sin\omega t\int_{-\infty}^{+\infty} d\omega\left[A_\omega\sin\omega t\int_0^\infty dT\left\{\frac{\omega T}{1+\omega^2T^2}\phi(T)\right\}\right]\right\}. \qquad \text{.....(A 7)}$$

The indentation-time function from Hertz's theory may be found as a numerically given function and the Fourier transform of this function may also be found numerically. This determines A_ω. Since the time scale is such that the maximum indentation occurs at $t=0$, A_ω is an even function and the real part of the velocity is equal to the total velocity, so that

$$-\int_{-\infty}^{+\infty} d\omega\,A_\omega\,\omega\sin\omega t=\frac{dx(t)}{dt}. \qquad \text{.....(A 8)}$$

The integration over t in (A 7) consists in evaluating

$$\int_{-\infty}^{+\infty} \frac{dx(t)}{dt} \sin \omega t \, dt = B_\omega \quad \text{(say)}. \qquad \ldots\ldots\text{(A 9)}$$

But
$$\frac{dx(t)}{dt} = \int_0^\infty b_\omega \sin \omega t \, d\omega, \qquad \ldots\ldots\text{(A 10)}$$

where
$$b_\omega = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{dx(t)}{dt} \sin \omega t \, dt$$

whence $\pi b_\omega = B_\omega$. Comparing equations (A 8) and (A 10) it is seen that $B_\omega/2\pi = -\omega A_\omega$. Thus

$$\overline{W} = 2\pi \int_{-\infty}^{+\infty} d\omega \int_0^\infty dT \omega A_\omega^2 \frac{\omega T}{1+\omega^2 T^2} \phi(T). \qquad \ldots\ldots\text{(A 11)}$$

In order to interpret equation (A 11) a simplifying approximation must be made. A_ω may be roughly represented by a step function and ωA_ω^2, which is a function having a single, sharp maximum, will thus approximate to a delta

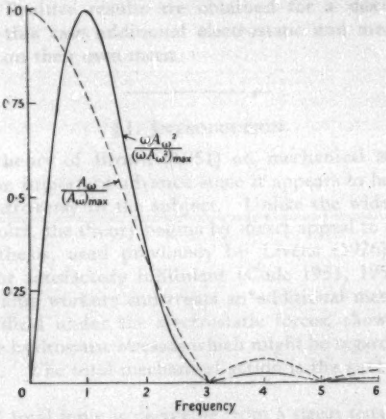


Figure 7. Fourier analysis of the pressure pulse. The form of the functions A_ω and ωA_ω^2. The unit of frequency is the reciprocal of twice the time of contact.

function. The form of ωA_ω^2 is shown in figure 7, and it is seen that the maximum occurs at $f' \simeq 1/2T_H$, where T_H is the time of contact calculated from Hertz's theory.

But $2\pi f' = \omega'$, so that $\omega' = \pi/T_H$. Since the delta function acts as a substitution operator such that

$$\int f(x)\delta(x-a)\,dx = f(a),$$

the integration over ω in (A 11) may be replaced by the substitution of ω' in the remainder of the formula. Thus

$$\frac{1}{2\pi} \overline{W}_{\omega'} \propto \int_0^\infty dT \frac{\omega' T}{1+\omega'^2 T^2} \phi(T) \qquad \ldots\ldots\text{(A 12)}$$

where $\overline{W}_{\omega'}$ is the energy loss per cycle occurring at a pulsatance ω'.

This equation is familiar as the relation between the imaginary part of the modulus and the relaxation-time spectrum. Therefore the experimental determination of $\overline{W}_{\omega'}$ is roughly equivalent to the determination of E_{ω}'', and the energy loss during impact may be attributed to loss at an 'equivalent frequency' f' of $1/2T_{\mathrm{II}}$. The above argument derives a value for the equivalent frequency to a first approximation only, but does give justification to the assumption of Jenckel and Klein.

When the losses due to internal friction are great the deformation will not follow the Hertz pattern. In any case the state of stress and strain in the region of contact is not uniform. In view of this the error involved in the approximation from equation (A 11) to (A 12) is probably not great in comparison with other errors.

Contributions to the Theory of Electrostatic Forces on Immersed Bodies

By R. CADE
University College of the West Indies, Jamaica

Communicated by R. Fürth; MS. received 15th February 1954

Abstract. Using the principles of Brown's theory of electrostatic forces on immersed bodies, an amendment is made to this theory to take into account, on a classical basis, the electrical and mechanical properties of the boundary of the body. A significant new feature is the inclusion of boundary-viscosity which, far from being negligible, is the all-important entity in the mechanics of the problem. Negative results are obtained for the case of a continuous-potential boundary, but they are such as to show the prior need for the deduction in a rigorous theory. Positive results are obtained for a discontinuous-potential boundary, and in this case additional electrostatic and mechanical properties appear, of interest on their own merit.

§ 1. Introduction

The recent theory of Brown (1951) on mechanical action in dielectrics constitutes an important advance since it appears to have largely resolved the long controversy in the subject. Unlike the widely accepted energy method of Helmholtz, the theory begins by direct appeal to the Poisson–Kelvin polarization hypothesis, used previously by Livens (1926) and Smith-White (1949) but without satisfactory fulfilment (Cade 1951, 1952). Brown avoids the errors of previous workers and treats an additional matter, the hydrostatic equilibrium of a fluid under the electrostatic forces, showing that there is a modification of the hydrostatic stresses which might be regarded as a 'secondary' mechanical action. The total mechanical action is the sum of the primary and secondary parts.

The density of total force is derivable from a stress tensor $T_{\mu\nu}$. It is found that (in Cartesian tensor notation)

$$T_{\mu\nu}=R_{\mu\nu}+(S_{\mu\nu}-S_{\mu\nu}^{*}) \qquad \ldots\ldots(1)$$

where, in terms of electric intensity E_μ and displacement D_μ,

$$R_{\mu\nu}=\frac{1}{4\pi}(D_\mu E_\nu-\tfrac{1}{2}\delta_{\mu\nu}E_\lambda E_\lambda). \qquad \ldots\ldots(2)$$

Here $\delta_{\mu\nu}$ is the fundamental second-rank tensor and $S_{\mu\nu}$ is related to the secondary force and has the value $S_{\mu\nu}^{*}$ at zero field. For a fluid at rest

$$S_{\mu\nu}=-p\delta_{\mu\nu}, \qquad \ldots\ldots(3)$$

and the scalar p is determined from the equation of equilibrium of the fluid under the primary force. The mechanical action on an immersed body is found by applying (1) at the outer surface of the body.

Brown is able to solve the equation of motion of a fluid under the assumption of static equilibrium to obtain general expressions for p, and hence $T_{\mu\nu}$. The result, unlike that for $R_{\mu\nu}$, depends upon properties of the fluid additional to the dielectric constant ϵ. In the case of a uniform, incompressible fluid in which D_μ is linear with respect to E_μ, which is the case of greatest physical importance since it is that to which most liquids approximate, it is found that $T_{\mu\nu}$ is the *Maxwell stress tensor*

$$(\epsilon/4\pi)(E_\mu E_\nu - \tfrac{1}{2}\delta_{\mu\nu}E_\lambda E_\lambda),$$

a fact quite remarkable in view of the long controversy in which Maxwell's assumption was generally considered obsolete. In the event of the fluid not satisfying the condition of incompressibility, the results obtained by Brown are entirely new. The value of $T_{\mu\nu}$ depends only upon the *local* properties of the fluid; in other words, if $T_{\mu\nu}$ is calculated at the surface S under the assumption of the dielectric having certain properties there, the result is unaffected if these properties do not obtain inside S. In particular, if the fluid dielectric contains an immersed rigid body surrounded infinitely closely by S, one might conclude with Brown that $T_{\mu\nu}$ gives the mechanical action on the body, although actually this final step involves an assumption.

In this paper we shall attempt an amendment of Brown's theory by the consideration on classical lines of phenomena at the boundary of an immersed body. It is customary in classical theory to treat such a boundary as the limit obtained when a thin layer in which dielectric constant changes sharply but continuously becomes of zero thickness. Now in all theories of mechanical action, including Brown's, the appropriate stresses are considered to be those in the fluid and at the boundary, which involves the tacit assumption that before the limit of a discontinuous boundary is reached the stresses acting on the solid are those at the side of the boundary layer remote from the solid. That such an assumption should be made is not obvious either on classical or molecular theory, for the boundary layer is also a region in which matter changes from fluid to solid, and the stresses should really be taken where solid is considered to begin. In Brown's theory the assumption of static equilibrium in the body of the fluid provides certain functional relationships representing *conditions of equilibrium*. The key to our treatment of the problem of the boundary layer is the fact that the abrupt change of electrical and mechanical quantities which takes place there may be such that *conditions of equilibrium are not satisfied*, and this possibility requires that the treatment be hydrodynamical.

§2. Formulation of the Problem

Consider a very thin transition layer in which certain quantities change sharply on going from fluid to solid. The faces of the boundary layer will be curved surfaces, and we introduce orthogonal curvilinear coordinates $x^\mu(\mu = 0, 1, 2)$ such that the surfaces in fluid and solid are given by $x^0 = c_f$ and $x^0 = c_s$, respectively, where c_f and c_s are constants and, for definiteness, $c_f > c_s$. Putting $c_f - c_s = a$, the limit of a boundary layer of zero thickness is the limit $a \to 0$. Associated with the coordinates is the metric $g_{\mu\nu}$, and it is necessary to discriminate between covariant and contravariant transformation properties. The only relationship between a vector v_μ say, in the new coordinates and its value v_μ' in the Cartesian coordinates is that provided by the transformation equations; v_μ has not, in general, the same physical meaning as v_μ'. Physical equivalence

can, if necessary, be obtained by introducing a unitary vector base $\{\mathbf{a}^\mu\}$ and its reciprocal $\{\mathbf{a}_\mu\}$, so that, if $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are the Cartesian unit vectors,

$$v_\mu \mathbf{a}_\mu = v_x'\mathbf{i} + v_y'\mathbf{j} + v_z'\mathbf{k},$$

and in the case of a contravariant vector

$$u^\mu \mathbf{a}^\mu = u_x'\mathbf{i} + u_y'\mathbf{j} + u_z'\mathbf{k}.$$

From the unitary vector bases unitary dyad bases may be formed in order to relate second-rank tensors in the same way. For convenience we shall use the same terminology to describe quantities in the new coordinates as in the Cartesian coordinates. For example, $v^\mu = dx^\mu/dt$ will be called the 'velocity', although the actual velocity, which we distinguish by the words 'physical velocity', is $(dx^\mu/dt)\mathbf{a}^\mu$.

The curve $x^1 = \text{constant}$, $x^2 = \text{constant}$ will intersect the surfaces $x^0 = c_s$, $x^0 = c_f$ at the points B_s, B_f, say, and we shall denote the respective values of a variable α at these points by α_s, α_f. The variables we have to deal with fall broadly into three types. If α is such that $\alpha_s - \alpha_f \to 0$ as $a \to 0$, it will be called 'slowly varying'. If $\alpha_s - \alpha_f$ tends to a finite, non-zero limit as $a \to 0$, α will be called 'quickly varying'. If $\alpha \to \infty$ as $a \to 0$ it will be called 'large'. Any function in the boundary layer is expressible in the form $f[x^1, x^2, (x^0 - c_s)/a, a]$. We make the physically reasonable assumption that all such functions can be expressed as series of integral powers of a, so that a quickly varying function can be expanded to

$$f_1[x^1, x^2, (x^0 - c_s)/a] + \sum_{n=1}^{\infty} a^n f_n.$$

The x^1-derivative of this is

$$f_1^{(x^1)}[x^1, x^2, (x^0 - c_s)/a] + \frac{\partial}{\partial x^1} \sum_{n=1}^{\infty} a^n f_n$$

and the x^0-derivative

$$a^{-1} f_1^{((x^0 - c_s)/a)}\left(x^1, x^2, \frac{x^0 - c_s}{a}\right) + \frac{\partial}{\partial x^0} \sum_{n=1}^{\infty} a^n f_n.$$

The first two expressions above are equal to $O(a^0)$, and the last to $O(a^{-1})$, as determined by the first term, which is arbitrarily larger than the second according to the smallness of a. A function may not depend upon the boundary layer, in which case it is expressible as $f(x^0, x^1, x^2)$ and, in view of the smallness of a,

$$f = f_s + a\left(\frac{x^0 - c_s}{a}\right)\left(\frac{\partial f}{\partial x^0}\right)_{x^0 = c_s} = \text{fn. of } x^1 \text{ and } x^2 + O(a).$$

In the manner we have introduced, all functions of x^0 pertaining to points in the boundary layer may be described in terms of orders of powers of a, and in the case of quickly varying functions and functions derivable from quickly varying functions by differential and integral operations with respect to x^0, differentiation with respect to x^1 and x^2 does not change the order, but differentiation with respect to x^0 decreases the order by one, while integration with respect to x^0 increases the order by one. The metric does not depend upon a, and therefore $g_{\mu\nu}$ and all related quantities, such as $g^{\mu\nu}$, g, $\{\mathbf{a}^\mu\}$, $\{\mathbf{a}_\mu\}$ and the Christoffel functions $[\mu\nu, \lambda]$ and $\{\mu\nu, \lambda\}$, are slowly varying. This property in the case of the unitary vectors shows that any variable has the same order as the associated 'physical' variable.

By the transformation of (1) and (2)

$$T^{\mu}{}_{\nu} = R^{\mu}{}_{\nu} + (S^{\mu}{}_{\nu} - S^{*\mu}{}_{\nu}), \qquad \ldots\ldots(4)$$

$$R^{\mu}{}_{\nu} = \frac{1}{4\pi}(D^{\mu}E_{\nu} - \tfrac{1}{2}\delta^{\mu}{}_{\nu}E_{\lambda}E^{\lambda}). \qquad \ldots\ldots(5)$$

As already explained, values of $S^{\mu}{}_{\nu}$ in the boundary layer are not necessarily those given by Brown's theory since the variation of quantities such as dielectric constant, charge density and mass density may be such as to preclude static equilibrium and require a steady state to be one of steady fluid motion (dynamic equilibrium). The components of $T^{\mu}{}_{\nu}$ which, taken at the real solid boundary, give the force on the immersed body are $T^{0}{}_{\nu}$, and if, according to (5) and the values obtained hydrodynamically for $S^{\mu}{}_{\nu}$, they are found to be quickly varying, a modification to Brown's results for the mechanical action on the body is necessary, for $T^{0}{}_{\nu}$ at the real solid boundary differ from $T_{f}{}^{0}{}_{\nu}$ at the fluid face of the boundary layer by amounts tending to non-zero limits as $a \to 0$. It might be discovered that velocity is quickly varying, in which case $v_{f}{}^{\mu}$ is non-zero in the limit $a \to 0$. Brown's results are then modified in an additional manner, for even if the conditions of equilibrium in the body of the fluid are satisfied, static equilibrium is impossible on account of the non-zero boundary velocity, and accordingly $T_{f}{}^{0}{}_{\nu}$ are modified. The conditions of equilibrium are, by their nature, necessary but not sufficient, and this would be a case of insufficiency *on electrostatic grounds*.

The principles upon which we shall treat the mechanics of the boundary layer will differ from Brown's in two respects. In the first place we shall not neglect viscosity and, indeed, it will emerge in the course of the calculations that *viscous stresses are the velocity functions of predominant importance*. One might anticipate this on other grounds; our concept of the boundary layer as a region in which the mobility of matter tends to zero on going towards solid suggests that even if viscosity is negligible in the body of the fluid this is not the case in the boundary layer at points close to the real solid boundary. We can, in fact, regard viscosity as a measure of mobility in the boundary layer, and we shall regard it as infinite at the real solid boundary.

The second point is that, since the boundary layer possesses direction, we must, for generality, assume viscous and dielectric anisotropy, and the latter implies that susceptibility and dielectric constant must be represented as second-rank tensors $\chi^{\mu}{}_{\nu}$ and $\epsilon^{\mu}{}_{\nu}$.

To introduce an anisotropic viscosity we assume that $S^{\mu}{}_{\nu}$ is related to $\delta^{\mu}{}_{\nu}$ and the velocity–strain tensor by

$$S^{\mu}{}_{\nu} = -p\delta^{\mu}{}_{\nu} + \alpha^{\mu}{}_{\nu}{}^{\varrho\sigma}\left(\frac{\partial v_{\varrho}}{\partial x^{\sigma}} - \{\rho\sigma, \lambda\}v_{\lambda}\right) \qquad \ldots\ldots(6)$$

where

$$-3p = S^{\mu}{}_{\mu}, \qquad \ldots\ldots(7)$$

and $\alpha^{\mu}{}_{\nu}{}^{\varrho\sigma}$ is a fourth-rank viscosity tensor. This theory is the natural extension of Brown's theory of polarized inviscid fluids and the St. Venant–Stokes theory of unpolarized isotropically viscous fluids (Lamb 1932, Weatherburn 1947), and is less restrictive than the latter theory because we are not assuming linearity with respect to $\partial v_{\mu}/\partial x^{\nu}$ (i.e. the boundary-layer fluid may be non-Newtonian). We shall assume only that $\alpha^{\mu}{}_{\nu}{}^{\varrho\sigma}$ tends to a finite, non-zero limit as $\partial v_{\mu}/\partial x^{\nu} \to 0$, which is fairly obvious, and is necessary for future discussions regarding the order of

$S^{\mu}{}_{\nu}$. Not all of the components of $\alpha^{\mu}{}_{\nu}{}^{\rho\sigma}$ are independent, on account of (7), but no extra restrictions are imposed by the form of $S^{\mu}{}_{\nu}$ because this tensor is not in general symmetric, as it is in the usual theory of stress. If the boundary layer were isotropic or inviscid, or had no velocity strains, D^{μ} would be parallel to E^{μ} and $S^{\mu}{}_{\nu}$ would be symmetric, but this is an important example of a situation with fluids which one might have thought rare, in which none of the three alternative conditions may hold.

Finally we briefly consider our classical model of the boundary layer in relation to the microscopic conception of a boundary as provided by experimental data, so that we may be sure that the correspondence is as complete as possible. At a fluid boundary there are forces of adhesion and surface tension; they are of the nature of short-range molecular forces, so that they are implicitly included in the mechanical stresses, contained in the tensor $S^{\mu}{}_{\nu}$. If we accepted the older view that it is only the surface *monolayer* of molecules that is modified by the surface forces, we should have to consider carefully whether these forces caused the boundary layer to be in effect rigidly bound to the solid, in which case it would be correct to take the stresses at the fluid face of the boundary layer in accordance with the previously mentioned implicit assumption of mechanical action theories, and our theory would be trivial. Even on this concept of a boundary, however, there seems little doubt that there are numerous cases of mobile monolayers, so that our theory would be applicable, but the whole question is rendered less serious by the great weight of recent experimental evidence in favour of the properties of liquids being modified in a 'surface zone' of considerable depth on the molecular scale (Henniker 1949). Henniker cites the experimental work of Deryagin on the viscosity of surface films. The results show viscosities of several times the normal values for the liquids. This is some of the decisive evidence in favour of the existence of the surface zone, and is in direct correspondence with a property we are assuming in our classical model. Still more recently M. and R. Freymann (1951) and Palmer *et al.* (1952) have investigated the dielecric constant of thin water films on solid surfaces and have obtained results intermediate between those for water and ice, a fact which conforms with the theory of the surface zone and with our idea of the quick variation of dielectric constant. The assumption of viscous and dielectric anistropy in the boundary layer is consistent with the known orientation of molecules and liquid crystals in the neighbourhood of a solid boundary. There seems to be no direct experimental evidence for such anisotropy, but optical anisotropy has been observed (Henniker 1949) and belongs to the strong evidence in support of the existence of the zone.

§ 3. Theory for a Continuous-Potential Boundary

A continuous-potential boundary is such that the potential is slowly varying, so that in the limit $a \to 0$

$$\phi_a = \phi_t. \qquad \ldots\ldots(8)$$

There may be boundary charge, which may be regarded as a volume distribution of density ρ in the boundary layer and ρ must be of order a^{-1} in order to give a finite, non-zero charge per unit area. We shall define surface density of charge at the point (x^0, x^1, x^2) by

$$\sigma = \int_{c_a}^{x^0} \rho \, dx^0; \qquad \ldots\ldots(9)$$

it is, of course, only the 'physical' surface density in the special case when x^0, x^1, and x^2 are Cartesian coordinates.

We generalize Brown's equation of motion so that it contains force on the charge and the more general stress system arising from viscosity. Its expression as a covariant vector equation is

$$\gamma\left(\frac{\partial v_\mu}{\partial t}+v^\nu v_{(\nu)\mu}\right)=\rho E_\mu+P^\nu E_{(\nu)\mu}-\gamma\frac{\partial\Omega}{\partial x^\mu}+S_{(\nu)}{}^\nu{}_\mu \qquad \ldots\ldots(10)$$

where P^μ is the polarization vector, γ the mass density and Ω the density of gravitational potential energy, and the bracketed suffixes correspond to covariant differentiation. The last term is, by (6),

$$S_{(\nu)}{}^\nu{}_\mu=(-p\delta^\nu{}_\mu)_{(\nu)}+\alpha_{(\nu)}{}^\nu{}_\mu{}^{\varrho\sigma}v_{(\sigma)\varrho},$$

and, since the covariant derivative of a fundamental tensor is zero,† this becomes

$$S_{(\nu)}{}^\nu{}_\mu=-\frac{\partial p}{\partial x^\mu}+\alpha_{(\nu)}{}^\nu{}_\mu{}^{\varrho\sigma}v_{(\sigma)\varrho}. \qquad \ldots\ldots(11)$$

We assume that steady conditions obtain, so that $\partial v_\mu/\partial t=0$, and treat (10) with a view to eliminating p. This can be done by taking the covariant derivative in the sense of an outer product, which gives a second-rank antisymmetric tensor equation. For any covariant vector A_μ

$$A_{(\nu)\mu}-A_{(\mu)\nu}=\frac{\partial A_\mu}{\partial x^\nu}-\frac{\partial A_\nu}{\partial x^\mu}, \qquad \ldots\ldots(12)$$

so that any vector like $\partial p/\partial x^\mu$, or

$$E_\mu=-\partial\phi/\partial x^\mu, \qquad \ldots\ldots(13)$$

disappears in this operation. We find for the 01-component of the new equation, using (11), (12) and (13)

$$\begin{aligned}\frac{\partial}{\partial x^0}(\gamma v^\nu v_{(\nu)1})-\frac{\partial}{\partial x^1}(\gamma v^\nu v_{(\nu)0})&=\left(\frac{\partial\rho}{\partial x^0}E_1-\frac{\partial\rho}{\partial x^1}E_0\right)+\left(\frac{\partial\gamma}{\partial x^1}\frac{\partial\Omega}{\partial x^0}-\frac{\partial\gamma}{\partial x^0}\frac{\partial\Omega}{\partial x^1}\right)\\&+\left(\frac{\partial P^\nu}{\partial x^0}\frac{\partial E_\nu}{\partial x^1}-\frac{\partial P^\nu}{\partial x^1}\frac{\partial E_\nu}{\partial x^0}\right)+\left(\frac{\partial}{\partial x^1}(\{\nu 0,\mu\}E_\mu P^\nu)-\frac{\partial}{\partial x^0}(\{\nu 1,\mu\}E_\mu P^\nu)\right)\\&+\frac{\partial}{\partial x^0}\alpha_{(\nu)}{}^\nu{}_1{}^{\varrho\sigma}v_{(\sigma)\varrho}-\frac{\partial}{\partial x^1}\alpha_{(\nu)}{}^\nu{}_0{}^{\varrho\sigma}v_{(\sigma)\varrho}. \qquad \ldots\ldots(14)\end{aligned}$$

Consider now the equation of continuity:

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^\mu}(\sqrt{g}\gamma v^\mu)=0. \qquad \ldots\ldots(15)$$

It follows from this equation that if v_1 and v_2 are quickly varying, v_0 is slowly varying and of the order of a, and, indeed, whatever the order of v_1 and v_2 in a, v_0 is of one order higher. Applying this fact to (14), we find that the functions of velocity of lowest order in a are the second x^0-derivatives of components of the viscosity tensor multiplied into $\partial v_1/\partial x^0$ and $\partial v_2/dx^0$, and it therefore seems justifiable to neglect all other velocity terms. We do this provisionally, the soundness of the step to be judged against the solution obtained, and the only surviving velocity term in (14) is

$$\frac{\partial^2}{(\partial x^0)^2}\left(\alpha^0{}_1{}^{10}\frac{\partial v_1}{\partial x^0}+\alpha^0{}_1{}^{20}\frac{\partial v_2}{\partial x^0}\right),=\frac{\partial^2 S^0{}_1}{(\partial x^0)^2} \qquad \ldots\ldots(16)$$

effectively, any error being of one order higher in a.

† For this, and other theorems of tensor analysis, see, for example, Weatherburn (1950).

The electric displacement satisfies the equation

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^\mu}(\sqrt{g}D^\mu)=4\pi\rho, \qquad \ldots\ldots(17)$$

which may be written

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^\mu}(\sqrt{g}\epsilon^{\mu\nu}E_\nu)=4\pi\rho. \qquad \ldots\ldots(18)$$

We know that $\rho=O(a^{-1})$ and that $\epsilon^{\mu\nu}$ is quickly varying, and since ρ and $\epsilon^{\mu\nu}$ are independent, it follows from this equation that E_0 is quickly varying although, on account of (8), E_1 and E_2 are slowly varying. Other quickly varying quantities are P^μ and γ, and we find that all terms in (14), other than the charge and velocity terms, are of order a^0 or a^{-1} according to whether the quickly varying quantities are differentiated with respect to x^0. The only terms not yet considered are the charge terms, and we see at once that $(\partial\rho/\partial x^1)E_0$ and $(\partial\rho/\partial x^0)E_1$ are of order a^{-1} and a^{-2} respectively. Collecting our information, we find that (14) may be reduced to the simple form

$$\frac{\partial^2 S^0{}_1}{(\partial x^0)^2}=-\frac{\partial\rho}{\partial x^0}E_1+O(a^{-1}). \qquad \ldots\ldots(19)$$

The integration of (19) can be carried out without difficulty. If matter is solid at the surface $x^0=c(x^1,x^2)$, we find that, at the real solid boundary,

$$S^0{}_1=S_f{}^0{}_1-\{\sigma(c)-\sigma(c_f)\}E_1+O(a), \qquad \ldots\ldots(20)$$

so that $S^0{}_1$, which is a tangential viscosity stress, is *quickly varying*.

The whole of the above analysis may be repeated for the 02-equation corresponding to (14), and we find that $S^0{}_2$ is given by (20) with the suffix 2 replacing 1.

Integration of (20) and the corresponding 02-equation, with the use of (6), will give v_1 and v_2, which we find, without doing the calculation, to be of order a. The fact that the velocity is non-zero is the complete justification of our original idea of the boundary layer not being in static equilibrium, but its slow variation shows that, in the limit $a\to 0$, there can be no violation of an assumption of the fluid as a whole being static or, more generally, following a given initial motion. The last statement means that (20) may be written

$$S^0{}_1=S_f{}^{*0}{}_1-(\sigma(c)-\sigma(c_f))E_1+O(a), \qquad \ldots\ldots(21)$$

and in the special case of zero applied field, which implies $\sigma=0$, even if this is an independent variable,

$$S^{*0}{}_1=S_f{}^{*0}{}_1. \qquad \ldots\ldots(22)$$

The tensor $R^\mu{}_\nu$, given by (5), may be manipulated with the use of (17), so that

$$\begin{aligned}\frac{\partial R^0{}_1}{\partial x^0}&=\rho E_1-\frac{E_1}{4\pi}\left(\frac{\partial D^1}{\partial x^1}+\frac{\partial D^2}{\partial x^2}\right)-\frac{E_1 D^\mu}{8\pi g}\frac{\partial g}{\partial x^\mu}+\frac{D^0}{4\pi}\frac{\partial E_1}{\partial x^0}\\&=\rho E_1+O(a^0). \qquad \ldots\ldots(23)\end{aligned}$$

Integrating this and taking the result at $x^0=c$, we obtain

$$R^0{}_1=R_f{}^0{}_1+\{\sigma(c)-\sigma(c_f)\}E_1+O(a). \qquad \ldots\ldots(24)$$

Combining (20), (22) and (24), and the corresponding 02-equations, in accordance with (4), and proceeding to the limit $a\to 0$, we find that

$$T^0{}_1=T_f{}^0{}_1, \qquad T^0{}_2=T_f{}^0{}_2. \qquad \ldots\ldots(25)$$

It remains for us to determine T^0_0, and for this purpose we make direct use of the x^0-component of (10) together with the x^0-derivative of R^0_0. By a calculation which is actually simpler, but which employs the same methods as we have used above, we then find that

$$T^0_0 = T^0_{f\,0}. \qquad \ldots\ldots(26)$$

The results we have obtained for S^0_r and v_μ show that our initial neglect of several of the velocity terms has been justified.

In conclusion, T^0_r and v_μ are slowly varying, so that effects at the boundary of an immersed body do not incur any modification to Brown's results for the electrostatic force on the body. Although the result is negative, the calculation from which it emerged is of significance, for it is found that, for a proper understanding of the electrostatic stresses on an immersed body, primary electrical forces must be considered in association with viscous forces. If we attempted to investigate the boundary layer stresses from an initial assumption of the fluid being in static equilibrium, the conclusion would be that T^0_r were modified. We find as a *deduction* that the fluid can be static in the limit $a \to 0$, but, even so, that there are hydrodynamical (viscosity) stresses, and these stresses just compensate changes in the primary force.

§4. Theory for a Discontinuous-Potential Boundary

A discontinuous-potential boundary is such that ϕ is quickly varying, so that, in the limit $a \to 0$,

$$\phi_e = \phi_f + \Phi. \qquad \ldots\ldots(27)$$

The potential difference Φ (a 'contact potential') is a necessary and sufficient condition for the existence of a surface distribution of dipole ('electric double layer') at the boundary. When the boundary is considered on classical lines as a thin boundary layer, the question arises as to whether the surface distribution of dipole should be regarded as a volume distribution of dipole (a 'permanent' polarization) or a volume distribution of charge. We find that our results depend upon what assumption we make, and this point requires examination, since Brown finds (1951) that his results hold equally if one replaces dielectric polarization by the Poisson charge distributions. The apparent discrepancy can be resolved by microscopic considerations, for, as discussed by Brown (private communication), the direct use of the Poisson distribution of volume charge for calculating the force on a volume element amounts to regarding the surface of the element as dissecting the molecular dipoles. In his theory, as one would expect, the difference in the primary force is exactly compensated by a difference in the secondary force, the latter, in this case, including the molecular binding forces. In our theory *relative motion of adjoining layers occurs*. It is clearly not the same for different molecules to move relatively as it is for charges constituting the same molecule to move relatively, and we therefore see that the two possibilities correspond to different physical circumstances, so that it is natural to expect different results. Actually, the two possibilities represent extremes of an infinity of possibilities corresponding to an infinite gradation of ratios of the two effects; while in Brown's theory all lead to the same results, in the present theory there will be a continuous gradation of results. We can, however, formulate our theory in general terms by assuming that the dipole sheet is made up partly of a charge

distribution and partly of a dipole distribution, the two parts being respectively contained in ρ and P^μ.†

An important difference from the case of Φ zero is that, in order for ϕ to be quickly varying, ρ and P^μ must be of order a^{-2} and a^{-1} respectively. A further difference is that since the dipole sheet is not under the control of the experimenter, there is an internal field even when there is no applied external field. It follows that (4) must be replaced by

$$T^\mu{}_\nu = (R^\mu{}_\nu - R^{*\mu}{}_\nu) + (S^\mu{}_\nu - S^{*\mu}{}_\nu), \qquad \ldots\ldots(28)$$

the asterisk, as always, denoting values at zero applied field.

Consider equation (14) as applied to the present situation. We may anticipate the conclusion which will be reached, to the effect that a number of the velocity terms which were previously neglected cannot be neglected now, and this complicates the problem of calculating the stresses. Our procedure will be to separate from (14) quantities of lowest order in a, which is a^{-3}, such that, apart from the factor a^{-3}, a only occurs in the quotient $(x^0 - c_s)/a$. These quantities form an independent equation with $S^\mu{}_\nu$ which is free of the additional velocity terms and from which the part of the velocity of lowest order in a can be calculated. The separation requires that we split variables into parts, which we distinguish by the upper suffixes [1] and [2], such that, in each case, the latter part is of one order higher in a than the former and is simply the first term of an expansion in powers of a, ($(x^0 - c_s)/a$ being, of course, an independent variable). For example, ρ, P^0 and E_0 have parts $\rho^{[1]}$, $P^{[1]0}$ and $E_0^{[1]}$ and $\rho^{[2]}$, $P^{[2]0}$ and $E_0^{[2]}$, the latter being of the special type described and *belonging exclusively to the dipole sheet*, and the metrical quantity g^{00} has parts $g^{[1]00}$ and $g^{[2]00}$, of which the latter is identical to g_s^{00}.

The 01-equation of order a^{-3} is

$$0 = \left(\frac{\partial \rho^{[2]}}{\partial x^1} E_0^{[2]} - \frac{\partial \rho^{[2]}}{\partial x^0} E_1^{[2]}\right) + \left(\frac{\partial P^{[2]0}}{\partial x^1}\frac{\partial E_0^{[2]}}{\partial x^0} - \frac{\partial P^{[2]0}}{\partial x^0}\frac{\partial E_0^{[2]}}{\partial x^1}\right)$$
$$- \frac{\partial^2}{(\partial x^0)^2}\left(\alpha^0{}_1{}^{10[2]}\frac{\partial v_1^{[2]}}{\partial x^0} + \alpha^0{}_1{}^{20[2]}\frac{\partial v_2^{[2]}}{\partial x^0}\right) + \{01,0\}^{[2]}\frac{\partial}{\partial x^0}(E_0^{[2]} P^{[2]0}). \qquad \ldots\ldots(29)$$

We manipulate the electrical terms with the aid of (17) and the definition

$$D^\mu = E^\mu + 4\pi P^\mu \qquad \ldots\ldots(30)$$

so as to obtain everything in terms of E_μ and D^μ, and integrate the result twice with respect to x^0. We thus obtain,

$$\alpha^0{}_1{}^{10[2]}\frac{\partial v_1^{[2]}}{\partial x^0} + \alpha^0{}_1{}^{20[2]}\frac{\partial v_2^{[2]}}{\partial x^0} = \frac{1}{4\pi}\int_{c_f}^{x^0}\left(\frac{\partial L^0}{c x^1}E_0 - \frac{\partial D^0}{\partial x^0}E_1 + \tfrac{1}{2}g^{00}\frac{\partial g_{00}}{\partial x^1}D^0 E_0\right)^{[2]} dx^0. \qquad (31)$$

We shall henceforth denote the right-hand side of this equation by A_{10}. We may treat the 02-equation corresponding to (29) in the same way and get an equation between $v_\mu^{[2]}$ and a quantity A_{20}, differing from (31) simply by the replacement of the suffix 1 by 2. Solving these equations for $\partial v_1^{[2]}/\partial x^0$ and $\partial v_2^{[2]}/\partial x^0$ and integrating again, we obtain for the velocity

$$\left.\begin{aligned} v_1^{[2]} &= \int_c^{x^0} \frac{\alpha^0{}_2{}^{20[2]}A_{10} - \alpha^0{}_1{}^{20[2]}A_{20}}{\alpha^0{}_2{}^{20[2]}\alpha^0{}_1{}^{10[2]} - \alpha^0{}_1{}^{20[2]}\alpha^0{}_2{}^{10[2]}}\,dx^0, \\ v_2^{[2]} &= \int_c^{x^0} \frac{\alpha^0{}_1{}^{10[2]}A_{20} - \alpha^0{}_2{}^{10[2]}A_{10}}{\alpha^0{}_2{}^{20[2]}\alpha^0{}_1{}^{10[2]} - \alpha^0{}_1{}^{20[2]}\alpha^0{}_2{}^{10[2]}}\,dx^0. \end{aligned}\right\} \qquad \ldots\ldots(32)$$

† We may still regard P^μ as related to E_μ by the susceptibility tensor $\chi^{\mu\nu}$.

We see at once that $v_1^{[2]}$ and $v_2^{[2]}$ are quickly varying. Formulae (32) are therefore important, not only for our subsequent calculations, but because this is a case where the body of the fluid connot be static even if Brown's conditions of equilibrium are satisfied. The quantities $v_{f1}^{[2]}$ and $v_{f2}^{[2]}$, obtainable from (32) simply by replacing the upper limits x^0 by c_f, are therefore required as boundary conditions in hydrodynamical deductions to determine the $T_f{}^0{}_\nu$, which are in general going to be different from the values given by Brown's theory.

We now turn to the question of calculating $T^0{}_1$ in the boundary layer, which, as in the case of $\Phi = 0$, is done by use of the 01-component of (14) and $R^0{}_1$. Since v_μ is of order a^0 we can sort out immediately the velocity terms which we previously neglected and cannot neglect now because they give quickly varying contributions to $S^0{}_1$. We can drop the suffix [2], having obtained formulae for the velocity, for the addition of $v_\mu^{[1]}$ only implies the addition of terms of order a to $S^0{}_1$, which therefore vanish in the limit $a \to 0$. Also, in our future treatment of electrical quantities, we shall not need a separation into parts with suffixes [1] and [2]. The additional velocity terms are

$$\frac{\partial}{\partial x^0} S_{(\mu)}{}^\mu{}_1 - \frac{\partial^2 S^0{}_1}{(\partial x^0)^2} - \frac{\partial^2 S^0{}_0}{\partial x^1 \partial x^0} = \frac{\partial}{c x^0}\left[\frac{\partial}{\partial x^\mu}\left(\alpha^\mu{}_1{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^\mu{}_1{}^{20}\frac{\partial v_2}{\partial x^0}\right)\right.$$
$$\left. - \{1\mu, \nu\}\left(\alpha^\mu{}_\nu{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^\mu{}_\nu{}^{20}\frac{\partial v_2}{\partial x^0}\right) + \{\nu\mu, \mu\}\left(\alpha^\nu{}_1{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^\nu{}_1{}^{20}\frac{\partial v_2}{\partial x^0}\right)\right]$$
$$- \frac{\partial^2}{(\partial x^0)^2}\left(\alpha^0{}_1{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^0{}_1{}^{20}\frac{\partial v_2}{\partial x^0}\right) - \frac{\partial^2}{\partial x^1 \partial x^0}\left(\alpha^0{}_0{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^0{}_0{}^{20}\frac{\partial v_2}{\partial x^0}\right). \quad \ldots\ldots(33)$$

We shall denote the right-hand side of this equation by $\partial\Sigma_{01}/\partial x^0$. We emphasize that $\partial\Sigma_{01}/\partial x^0$ can, in principle, be calculated from (32); $\partial^2 S^0{}_1/(\partial x^0)^2$ cannot, of course, because three successive differentiations with respect to x^0 are involved, so that $v_\mu^{[1]}$ is of importance, and this cannot be calculated *a priori*.

The 01-component of (14) may now be written

$$\frac{\partial^2 S^0{}_1}{(\partial x^0)^2} = \left(\frac{\partial \rho}{\partial x^1} E_0 - \frac{\partial \rho}{c x^0} E_1\right) + \left(\frac{\partial P^\nu}{\partial x^1}\frac{\partial E_\nu}{\partial x^0} - \frac{\partial P^\nu}{\partial x^0}\frac{\partial E_\nu}{\partial x^1}\right)$$
$$- \left[\frac{\partial}{\partial x^1}(\{\nu 0, \mu\} E_\mu P^\nu) - \frac{\partial}{\partial x^0}(\{\nu 1, \mu\} E_\mu P^\nu)\right] - \frac{\partial \Sigma_{01}}{\partial x^0} \quad \ldots\ldots(34)$$

where all quantities which we already know to be of order a^{-1} or higher have been dropped. Consider the angular momentum equation of a volume element. Since the ratio of angular momentum to volume tends to zero as the volume tends to zero, we have, as in the usual theory of stress,

$$T_{\mu\nu} - T_{\nu\mu} = 0 \quad \ldots\ldots(35)$$

or, in terms of couple density,

$$(S_{\mu\nu} - S_{\nu\mu}) + (P_\mu E_\nu - P_\nu E_\mu) = 0. \quad \ldots\ldots(36)$$

Since $S^\mu{}_\nu$ and E_0 are of order a^{-1} and E_1 and E_2 are of order a^0, this equation shows that P^1 and P^2 *cannot be of order lower than* a^0. This is equivalent to the statement that to order a^{-1} in E_μ and P^μ, *one principal axis of* $\chi^{\mu\nu}$ *is perpendicular to the boundary*. We form the sum of (34) with the second x^0-derivative of $R^0{}_1$ and reduce the whole expression to its simplest possible form in terms of E_μ, P^μ and

D^μ by some tedious manipulation in which we use the above restriction on P^μ. Integrating the result twice, applying (28) and proceeding to the limit $a \to 0$, we obtain finally

$$\left.\begin{aligned} T^0{}_1 &= T^0_{f\,1} + \lim_{a\to 0}\int_{c_f}^{c} (I_{01} - I^*_{01})\,dx^0, \\ I_{01} &= \frac{1}{4\pi}\frac{\partial}{\partial x^1}(E_0 D^0) + \frac{E_0 D^0}{8\pi} g^{00}\frac{\partial g_{00}}{\partial x^1} - \frac{E_1 D^0}{8\pi}\frac{1}{g}\frac{\partial g}{\partial x^0} \\ &\quad + \tfrac{1}{2}(E_1 P^0 g^{11} - E_0 P^1 g^{00})\frac{\partial g_{11}}{\partial x^0} \\ &\quad - \Sigma_{01} + \frac{1}{8\pi}\int_{c_f}^{x^0}\frac{\partial}{\partial x^1}\left[E_0 D^0 \frac{\partial}{\partial x^0}\log(g_{11}g_{22})\right]dx^0. \end{aligned}\right\} \quad \ldots\ldots(37)$$

A similar analysis applied to the 02-component of (14) gives $T^0{}_2$, the formula for which differs from (37) simply by the replacement of the suffix 1 by 2. The quantity Σ_{02} which the result contains is related to Σ_{01} by the interchange of the suffixes 1 and 2 in (33).

To obtain $T^0{}_0$ we work directly with the 0-component of (10), as in the case of Φ zero. As in the calculation of $T^0{}_1$, we have to include velocity terms which we were able to omit with Φ zero, and these are

$$\begin{aligned} S_{(\mu)}{}^\mu{}_0 - \frac{\partial S^0{}_0}{\partial x^0} &= \frac{\partial}{\partial x^\mu}\left(\alpha^\mu{}_0{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^\mu{}_0{}^{20}\frac{\partial v_2}{\partial x^0}\right) \\ &\quad - \{0\mu, \nu\}\left(\alpha^\mu{}_\nu{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^\mu{}_\nu{}^{20}\frac{\partial v_2}{\partial x^0}\right) + \{\nu\mu, \mu\}\left(\alpha^\nu{}_0{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^\nu{}_0{}^{20}\frac{\partial v_2}{\partial x^0}\right) \\ &\quad - \frac{\partial}{\partial x^0}\left(\alpha^0{}_0{}^{10}\frac{\partial v_1}{\partial x^0} + \alpha^0{}_0{}^{20}\frac{\partial v_2}{\partial x^0}\right). \qquad \ldots\ldots(38) \end{aligned}$$

We shall denote the right-hand quantity by Σ_{00}, and bearing in mind that Σ_{00} can, in principle, be calculated from (32), we see that the 0-component of (10) is an equation for determining $S^0{}_0$. Calculation of $T^0{}_0$ by the same methods as we have already used leads to the result

$$\left.\begin{aligned} T^0{}_0 &= T^0_{f\,0} + \lim_{a\to 0}\int_{c_f}^{c} (I_{00} - I^*_{00})\,dx^0, \\ I_{00} &= -\frac{E_0 D^0}{8\pi}\frac{\partial}{\partial x^0}\log(g_{11}g_{22}) - \frac{1}{4\pi\sqrt{g}}\left(\frac{\partial}{\partial x^1}(E_0 D^1\sqrt{g})\right. \\ &\quad \left. + \frac{\partial}{\partial x^2}(E_0 D^2\sqrt{g})\right) + \tfrac{1}{2}(E_0 P^1 g^{00} - E_1 P^0 g^{11})\frac{\partial g_{00}}{\partial x^1} \\ &\quad + \tfrac{1}{2}(E_0 P^2 g^{00} - E_2 P^0 g^{22})\frac{\partial g_{00}}{\partial x^2} - \Sigma_{00}. \end{aligned}\right\} \quad \ldots\ldots(39)$$

We have therefore found that, when there is a dipole sheet at the boundary of an immersed body, Brown's results for the mechanical action on the body are modified from both possible causes, the existence of a discontinuity of velocity at the boundary and a quick variation of the $T^0{}_\nu$ in the infinitely thin boundary layer. The quick variation of the $T^0{}_\nu$ is given explicitly by (37) and (39). The effect of the discontinuity of velocity is a modification of the stresses in the body of the fluid, and must be calculated on the basis of the hydrodynamical conditions obtaining in individual cases. The operative condition is the boundary condition

provided by (32). Formulae (32), (37) and (39) therefore represent the complete results for the mechanical action on the body.

§ 5. Conditions upon the Dipole Sheet

We discovered in the course of the calculations that P^1 and P^2 are of order a^0. Since $P^{[2]\mu}$ of lower order must belong exclusively to the dipole sheet, this represents a condition imposed electrically and hydrodynamically upon the dipole sheet. We shall now show that our results lead to some more conditions.

Consider a finite immersed body and any closed surface $x^0 = \text{constant}$ in the boundary layer. The total stress at the surface, rather than the difference between the values when there is and when there is not an applied field, is given formally by $R^0{}_\nu + S^0{}_\nu$, obtainable from (37), the corresponding 02-formula and (39) simply by dropping the quantities with asterisks. The fact that E_0 and D^0 have parts $E_0^{[2]}$ and $D^{[2]0}$ of order a^{-1} means that certain parts of the stresses are infinite in the limit $a \to 0$ and it follows that the total force on the matter inside the surface may be infinite. This, apart from being physically undesirable, involves a mechanical inconsistency, however, as will now be shown. Since the $R_f{}^\mu{}_\nu + S_f{}^\mu{}_\nu$ are finite, if the $R^\mu{}_\nu + S^\mu{}_\nu$ are infinite and provide an infinite total force, there is an infinite total force of reaction on the part of the boundary layer between the arbitrary surface and the surface $x^0 = c_f$, and this may be equated to the total force of inertia. To consider the force of inertia we use Cartesian coordinates for a moment, for it is ultimately only with these that corresponding components of force at different parts of the boundary may be added algebraically to give a total force. We find that the acceleration is of order a^{-1}†, so that since $\gamma = \mathrm{O}(a^0)$, the total inertial force is finite. A similar problem arises regarding the torque; we find that the total torque on the part of the boundary layer may be of order a^{-1}, and yet the total rate of change of angular momentum is of order a^0.

We therefore have a contradiction, and the only way of resolving the difficulty is to find some term in the equation of motion which could take a large value and which, for some reason or other, we omitted. The only such term is the part $\partial v_\mu/\partial t$ of the acceleration, which we dropped following our assumption of dynamic equilibrium. We conclude that *unless the force and torque of order a^{-1} are zero, dynamic equilibrium is impossible.* Accordingly, an assumption of dynamic equilibrium requires that we *equate* these quantities to zero, and since the only field variables in them are $E_0^{[2]}$ and $D^{[2]0}$, belonging exclusively to the dipole sheet, the equations are *conditions of equilibrium for the dipole sheet.*

Consider first the force. To find the stress at the arbitrary surface $x^0 = \text{constant}$ we replace the upper limits of integration c by x^0 in (37), the corresponding 02-formula and (39). Inner multiplication of these equations by $\{\mathbf{a}_\mu\}$ and integration over the surface then gives the total force. Using the arbitrariness of the above-mentioned upper limit and retaining only the terms of the order we are concerned with, the required condition is found to be

$$\int_S \left\{ -\mathbf{a}_0 \tfrac{1}{2} E_0 D^0 \frac{\partial}{\partial x^0} \log(g_{11} g_{22}) + \mathbf{a}_1 \left(\tfrac{1}{2} g^{00} \frac{\partial g_{00}}{\partial x^1} E_0 D^0 + \frac{\partial}{\partial x^1}(E_0 D^0) \right) \right.$$
$$\left. + \mathbf{a}_2 \left(\tfrac{1}{2} g^{00} \frac{\partial g_{00}}{\partial x^2} E_0 D^0 + \frac{\partial}{\partial x^2}(E_0 D^0) \right) \right\}^{[2]} dS_0 = 0, \qquad \ldots\ldots (40)$$

† This is in peculiar contrast with the order a^0 obtained with our particular choice of curvilinear coordinates, but it is differences of this type which make the use of such co-ordinates mathematically advantageous.

where dS_r is the covariant surface-element vector. By consideration of the torque in the same manner, we obtain

$$\int_S \left\{ [\mathbf{a}_0, \mathbf{r}] \tfrac{1}{2} E_0 D^0 \frac{\partial}{\partial x^0} \log (g_{11} g_{22}) + [\mathbf{r}, \mathbf{a}_1] \left(\tfrac{1}{2} g^{00} \frac{\partial g_{00}}{\partial x^1} E_0 D^0 + \frac{\partial}{\partial x^1} (E_0 D^0) \right) \right.$$
$$\left. + [\mathbf{r}, \mathbf{a}_2] \left(\tfrac{1}{2} g^{00} \frac{\partial g_{00}}{\partial x^2} E_0 D^0 + \frac{\partial}{\partial x^2} (E_0 D^0) \right) \right\}^{[2]} dS_0 = 0 \qquad \ldots\ldots (41)$$

where $\mathbf{r}$ is the displacement vector from the point about which the moment is taken.

Yet other conditions are imposed on the dipole sheet by the fact that $v_\mu^{[2]}$, given by (32), must be zero when there is no applied field. If this was not the case, the energy dissipated in the dipole sheet would have to come ultimately from the dipole sheet itself, which would contradict the second law of thermodynamics. From (31) and (32) the conditions are

$$\frac{\partial D^{*[2]0}}{\partial x^\mu} E_0^{*[2]} - \frac{\partial D^{*[2]0}}{\partial x^0} E_\mu^{*[2]} + \tfrac{1}{2} g^{[2]00} \frac{\partial g_{00}^{[2]}}{\partial x^\mu} D^{*[2]0} E_0^{*[2]} = 0 \quad (\mu = 1, 2). \qquad \ldots\ldots (42)$$

We may regard these as conditions of equilibrium, for if they were not satisfied (say at the instant an applied field was switched off) there would, presumably, be a transient adjustment to a state in which they were.

Further discussion of the conditions we have found lies beyond the scope of this paper, but the reader can easily verify that they can be simultaneously satisfied by dipole sheets of simple structure surrounding surfaces of simple shape.

§ 6. Discussion

Collecting our results, we may state them generally as follows. If the potential at the boundary of an immersed body is continuous, no modification to Brown's results for the mechanical action on the body arises through phenomena at the boundary, treated classically as the limit of zero thickness of a thin boundary layer. This result is only achieved through the rigour of our deduction, in which we recognize that boundary viscosity, far from being negligible, is the significant entity in the hydrodynamical part of the problem. If there is a discontinuity of potential, which implies the existence of a surface distribution of dipole, Brown's results are modified, because of discontinuity both of velocity and of the stresses.

Two points of principle, of interest apart from their significance to the theory of electrostatic forces on immersed bodies, have emerged in the course of the deductions:

(i) Brown's conditions of equilibrium are insufficient *on pure electrostatic grounds* if there is a discontinuity of potential at a solid boundary;

(ii) a dipole sheet in viscous fluid must obey certain conditions of its own.

The integrals in (32), (37) and (39) cannot be evaluated without a detailed knowledge of the structure of the boundary layer, which suggests that, to obtain numerically definite results, one would be obliged to consider the theory in relation to molecular models. The following question then arises: since in reality a has a fixed value, were we justified in treating parts of equations of different orders in a as independent? The premise of this question is of course arguable; one might possibly advance statistical arguments to show that, due to the agitation of molecules, a has not an absolutely fixed value, but can vary between narrow limits, and such a conclusion would suffice to justify our procedure. But in any

case, the question is not vital; we could have managed without the assumption and just treated parts of equations on the basis of their relative largeness. Certain conclusions would have been slightly different; for example, we need not have asserted that the condition of equilibrium (40) holds exactly. It would be inaccurate by a relatively very small quantity which would, in fact, be quite negligible in view of the extreme smallness of a. We made the assumption for two reasons: in the first place it is consistent with a purely classical theory, and secondly, in the author's opinion, it makes the theory mathematically simpler.

Motion of a fluid relative to a fixed solid boundary when there is a dipole sheet and applied field, such as our theory predicts, is well known experimentally in the field of colloid science; it belongs to the class of effects called electrokinetic phenomena and is usually known as electrosmosis. We have therefore given, in an incidental manner, a theory of electrosmosis, and it is of some importance to consider our treatment in relation to previous theories of the phenomenon. The problem was first treated by Helmholtz, subsequently by Lamb, Smoluchowski, Perrin and Debye–Hückel (the literature is cited, for example, by Thomas (1934)) and more recently by Henry (1931). The earlier theories were rather crude. The dipole sheet was regarded as a double layer of surface charges forming a 'parallel-plate condenser'. The view subsequently adopted was that of the 'diffuse double layer' in which one layer was a surface charge rigidly bound to the solid, while the other was a thin distribution of volume charge in the fluid and becoming negligible only at an appreciable distance (on the molecular scale) from the boundary. This model is used by Henry, whose theory is hydrodynamically rigorous within the framework of the assumptions, but is not free of mistakes in the electrostatics.†

Previous theories, in particular that of Henry, have some correspondence with the present theory, but the correspondence is not as close as would appear at first sight. Henry studies hydrodynamical problems involving dipole sheets in viscous fluids, but he regards the solid as having a definite, sharp boundary, with stresses to be taken outside the boundary as in previous theories of forces on immersed bodies. The whole essence of the present theory is to regard stresses as acting inside the boundary, which is regarded as the limit of a finite boundary layer, and, as explained, we are actually not obliged to proceed to the limit. It would appear that, from the point of view of electrokinetic theory equally with that of electrostatic forces, the approach we have adopted is the correct one, and this view is enhanced by the fact that a dipole sheet is in reality a surface phenomenon physically inseparable from other surface phenomena such as the quick variation of dielectric constant and viscosity.‡ The whole position is that the theory of electrokinesis is an inseparable part of the theory of electrostatic forces on immersed bodies; one cannot consider either without the other.

The usual formula for electrosmosis is that given by Smoluchowski and verified by Henry (1931):

$$v_{t\mu} = \frac{\epsilon_f \Phi E_\mu}{4\pi\alpha_f} \quad (\mu = 1, 2). \qquad \ldots\ldots(43)$$

† No account is taken of force due to the polarization of the fluid in the equations of motion, nor due to the polarization of the solid in the calculation of the pure electrical force.

‡ It is well known experimentally, through the phenomenon of electroviscosity (Elton 1948), that there is a correlation between viscosity and a dipole sheet of a more special type than that obtaining generally in the surface zone.

The present theory will give (43) under the following assumptions, which are in close correspondence with those of previous theories: (i) the surface is plane, cylindrical or spherical†, (ii) one layer of the charges constituting the dipole sheet is a surface charge lying entirely in solid, (iii) the dipole sheet consists solely of a charge distribution, (iv) $\rho^{[2]}$ is a function of x^0 only, (v) for $x^0 > c$ the fluid is uniform and isotropic with the same properties as for $x^0 > c_f$, so that

$$\epsilon^{\mu\nu} = \epsilon_f, \quad \alpha^0{}_1{}^{10} = \alpha^0{}_2{}^{20} = \alpha_f, \quad \alpha^0{}_1{}^{20} = \alpha^0{}_2{}^{10} = 0.$$

It seems to the author that it is *only* under these assumptions that (43) is valid, and this is an important point since electrokinetic theory is sometimes considered to be well verified by experiment. It appears, however, that experimental tests have always involved comparing the formula for elecrophoresis with that for electrosmosis, *the latter being regarded as basic* (cf. Henry 1931, Thomas 1934). Clearly there is room for considerable theoretical and experimental investigation.

Returning to the main question of the stresses on an immersed body, we consider some specific experiments which have revealed discrepancies probably traceable to mechanical action theory. Fürth (1924 a, b, 1927) calculated the couple on a conducting spheroid in a conducting fluid and influenced by a uniform electric field, and subsequently used the result as the basis of an experimental method for determining the dielectric constants of dilute aqueous solutions. The experimental results of Fürth and his followers (e.g. Lin 1936) showed, broadly speaking, two discrepancies. Firstly, the values of the dielecric constants differed appreciably from known or expected values, and secondly, an evidently spurious frequency dependence at low frequencies was found. In all experiments alternating fields were used, and it is easily seen, upon examination of the relevant formulae, that there will, in general, be a quick variation of the stresses in the boundary layer provided one or more of the relevant quantities is dependent upon the field strength, and also a modification of $T_f^{\mu}{}_{\nu}$ due to electrokinesis in the body of the fluid if one of the variable quantities is Φ. It is probable that Φ is field-dependent in conducting fluids, where it is contributed to by ionic migration, and it is even more likely that quantities such as the parts of $\rho^{[1]}$ and $P^{[1]\mu}$ belonging to the dipole sheet have this characteristic, for the energy associated with them is of a lower order. If there is a modification due to a variation of Φ, say, it is to be expected that the modification will depend upon the concentration of an aqueous solution, since this will determine the conditions of ionic migration, and, in fact, the most conspicuous discrepancies are with regard to the change of the couple with respect to concentration (Lin 1936). On similar grounds one might also explain the observed frequency dependence of the couple. If the mean value of Φ depends upon ionic migration in, say, a half-cycle of the field, and the ionic migration depends upon the duration of the half-cycle, then Φ is frequency dependent. We could discuss at length the various possibilities which could, on the basis of our theory, contribute to a modification of the stresses, but all such suggestions would, like the above, be based purely on conjecture. There is little upon which to judge the value of such guesses, save for some rather significant work of Fischer and Schaffeld (1936), who discovered that the frequency dependence of the couple depended upon *surface contamination* of the spheroid. This is

† It will be noticed that there is a curvature effect in (31) which is non-zero for surfaces other than the plane, cylinder and sphere. This was not obtained in the earlier theories, evidently because surfaces of general curvature were not considered.

good *prima facie* evidence in support of the ideas we propose, since the properties of dipole sheets are known to be sensitive to changes in the nature of surfaces, and it certainly suggests the discrepancies being traceable to surface phenomena.

Acknowledgments

The basic idea of this work evolved to a considerable extent from a suggestion made some time ago by Dr. R. Fürth of Birkbeck College, University of London, in connection with some other problems in electrostatics. I am indebted to Dr. Fürth for this, and for criticisms of an earlier version of the paper which led to substantial improvements. In the latter connection I acknowledge in particular his pointing out the relationship between the present theory and the theory of electrokinesis. I am also grateful to Dr. W. F. Brown, Jr., for a protracted discussion by correspondence which helped me to understand some of the finer points of his theory.

References

Brown, W. F., Jr., 1951, *Amer. J. Phys.*, **19,** 290, 333; 1953, *Rev. Mod. Phys.*, **25,** 131.
Cade, R., 1951, *Proc. Phys. Soc.* A, **64,** 665; 1952, *Ibid*, **65,** 287.
Elton, G. A. H., 1948, *Proc. Roy. Soc.* A, **194,** 259, 275.
Fischer, G., and Schaffeld, W. R., 1936, *Ann. Phys., Lpz.*, **25,** 450.
Freymann, M. and R., 1951, *C. R. Acad. Sci., Paris*, **232,** 401.
Fürth, R., 1924 a, *Z. Phys.*, **22,** 98; 1924 b, *Phys. Z.*, **25,** 676; 1927, *Z. Phys.*, **44,** 256.
Henniker, J. C., 1949, *Rev. Mod. Phys.*, **21,** 322.
Henry, D. C., 1931, *Proc. Roy. Soc.* A, **135,** 106.
Lamb, H., 1932, *Hydrodynamics* (Cambridge : University Press).
Lin, T., 1936, *Ann. Phys., Lpz.*, **26,** 495.
Livens, G. H., 1926, *Theory of Electricity* (Cambridge : University Press).
Palmer, L. S., Cunliffe, A., and Hough, J. M., 1952, *Nature, Lond.*, **170,** 796.
Smith-White, W. B., 1949, *Phil. Mag.*, **40,** 466.
Thomas, A. W., 1934, *Colloid Chemistry* (New York : McGraw-Hill).
Weatherburn, C. E., 1947, *Advanced Vector Analysis* (London : G. Bell); 1950, *Riemannian Geometry and the Tensor Calculus* (Cambridge : University Press).

Theory of Electron Beam Loading in Linear Accelerators

By G. SAXON
Research Department, Metropolitan-Vickers Electrical Co. Ltd., Manchester

Communicated by B. G. Churcher; MS. received 28th January 1954, and in final form 20th May 1954

Abstract. An analysis is made of the waveguide circuit of a linear accelerator of electrons employing radio-frequency power feedback to show how the power flowing into such an accelerator varies with the beam loading. The power relationships developed may be used in conjunction with formulae for the energy gain and beam power output of a length of accelerator waveguide to calculate the performance to be expected under feedback conditions as the beam current is varied.

If a bunching section forms part of an accelerator, it will have different characteristics from the rest of the accelerator and the analysis must be modified to take this into account. This applies particularly under the variable power input conditions occurring when feedback is used.

Measurements made on two linear accelerators of the variation of energy with beam current have been compared with the theoretically expected performance figures, and a reasonable agreement has been found.

§ 1. Introduction

In the design of linear electron accelerators the problem of achieving high energy gains in a short length with the radio-frequency power sources at present available has been overcome by the employment of power feedback, first suggested by Shersby-Harvie and Mullett (1949) and incorporated in the design of the 8 Mev linear accelerator now operating at Hammersmith Hospital (Miller 1953 a, b, 1954).

A beam-loading theory for accelerators employing feedback has been given by Johnsen (1951), but owing to certain simplifying assumptions his formulae give optimistic ideas of accelerator performances under feedback conditions. One of these assumptions is that power build-up, that is, the ratio of the accelerator input power to the source power in the steady state, remains constant with variation of loading. The analysis given below shows how the power build-up varies with the losses in the accelerator–feedback loop and hence with the power absorbed by the high energy electron beam. Later in the paper formulae are developed for the energy gain and beam power of any accelerator in terms of the current loading, and the power relationships arising from the analysis of the feedback circuit are applied to show how these formulae should be modified when feedback is used.

The inclusion of a bunching section in a length of accelerator waveguide makes the analysis of its behaviour under variable loading conditions rather more awkward. Due to its special characteristics, it is usually misleading to neglect its effect and assume a uniform guide. A relatively simple and, it is hoped,

adequate method of taking its properties into account in the analysis is given in §§ 3 and 5.

Waveguide design parameters are discussed in § 4, which gives data on which later calculations are based. These calculations include the comparison of theoretical and measured performances of two accelerators designed for x-ray therapy equipment.

Certain important factors in the design of linear accelerators, such as frequency and dimensional tolerances, are not discussed here since these questions have been dealt with adequately in previous publications (for example, Shersby-Harvie 1948). Such considerations should be taken in conjunction with the theory given here, which is intended to facilitate the calculation of expected performance of any accelerator waveguide with or without power feedback.

§ 2. Properties of the Feedback Circuit

The basic waveguide circuit is shown in figure 1 (*a*). The residual radio-frequency power from the accelerator is fed back, suitably phased, into a waveguide bridge where it combines with the input power and re-circulates. When a steady state is reached the power flow into the accelerator is greater than that from the source by a factor dependent only on the properties of the bridge and the attenuation of the accelerator-feedback loop. The power flow is, for instance, doubled for the particular case of a unity ratio bridge and a loop attenuation of 3 dB.

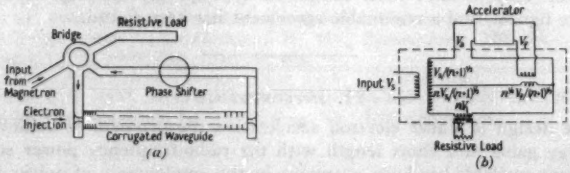


Figure 1. (*a*) Basic waveguide circuit of accelerator with feedback; (*b*) equivalent circuit.

The bridge ratio may be defined by a number n which is the ratio of the powers it is intended to combine. Thus if two conjugate arms of the bridge are fed with powers nW and W respectively, then, if the phasing is correct, one of the other arms will have a power flow of $(n+1)W$ and the remaining arm zero power flow. Standard waveguide bridges in which $n=1$ are the 'magic-tee' and the 'rat-race' with arms all of equal dimensions. For useful design information and an alternative theoretical approach to that presented below a memorandum by Mullett and Loach (1949) is recommended.

One equivalent circuit for a waveguide bridge is shown in figure 1 (*b*), in which the four arms have equal impedances. The values for the transformer ratios follow from insertion of the appropriate conjugate and combining properties of the bridge into the circuit equations. Let V_a, V_f and V_l be the instantaneous steady state voltages in the accelerator, feedback and resistive load arms of the bridge respectively, and V_s be the applied voltage from the source. Convenient expressions may be obtained if we employ a variable $x=|V_a|/|V_f|$ so that the attenuation of the accelerator-feedback loop is $10\log_{10}x^2$ dB or $\ln x$ nepers.

Making the appropriate vector summations, we get

$$\frac{V_a^2}{V_s^2} = \frac{W_a}{W_s} = \frac{x^2}{(n+1)x^2 - 2\{n(n+1)\}^{1/2} x\cos\phi + n}$$

$$\frac{V_l^2}{V_s^2} = \frac{W_l}{W_s} = \frac{(n+1)/n - 2\{(n+1)/n\}^{1/2} x\cos\phi + x^2}{n - 2\{n(n+1)\}^{1/2} x\cos\phi + (n+1)x^2}$$

where ϕ is the phase of the voltage in the feedback arm of the bridge relative to that in the accelerator arm. W_a and W_l refer to the power flows into the accelerator and resistive load respectively, whilst W_s is the power supplied from the source.

When the phase is correctly adjusted so that $\phi = 0$

$$\frac{W_a}{W_s} = \frac{x^2}{\{(n+1)^{1/2}x - n^{1/2}\}^2} \qquad \ldots\ldots(1)$$

$$\frac{W_l}{W_s} = \frac{[\{(n+1)/n\}^{1/2} - x]^2}{\{(n+1)^{1/2}x - n^{1/2}\}^2} \qquad \ldots\ldots(2)$$

These power ratios are plotted in figure 2 for the case $n=1$ as a function of the loop attenuation. It may be seen that when the latter is 3 dB ($x=\sqrt{2}$) then the power build-up ratio is 2 and no power goes into the resistive load. For any other particular value of the attenuation, if a bridge of correct ratio n were used such that $|V_a|^2/|V_f|^2 = (n+1)/n = x^2$, then the maximum build-up power ratio would be

$$W_a/W_s = n+1 = x^2/(x^2-1) \qquad \ldots\ldots(3)$$

and, under these circumstances, the power in the resistive load would, of course, be zero.

The build-up ratio given by the above expression (3) is shown in figure 2.

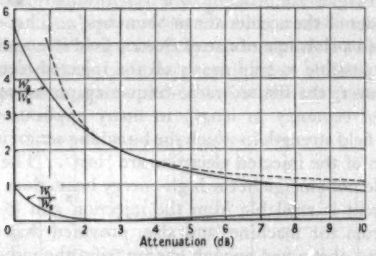


Figure 2. Properties of the feedback circuit. The broken line gives the maximum build-up obtainable for any given value of attenuation by using a bridge of correct ratio.

Comparison with the curve of equation (1) shows that a unity ratio bridge may be used over a wide range of attenuation without serious loss of power build-up. This is particularly fortunate in view of the fact that the attenuation of an accelerator may vary considerably with beam loading. The nature of the build-up action in the unity ratio bridge circuit may be considered as follows. A voltage V_s applied to the bridge initially produces a voltage $V_s/\sqrt{2}$ in both accelerator and resistive load arms. After one transit of the accelerator a voltage $V_s/x\sqrt{2}$

appears in the feedback arm, and, if the phase is correct, we shall now obtain voltages $V_s\{1+1/x\sqrt{2}\}/\sqrt{2}$ and $V_s\{1-1/x\sqrt{2}\}/\sqrt{2}$ in the accelerator and resistive load arms respectively. Further transits of the system lead to the series

$$V_a = \frac{V_s}{\sqrt{2}}\left(1+\frac{1}{x\sqrt{2}}+\frac{1}{(x\sqrt{2})^2}+\frac{1}{(x\sqrt{2})^3}+\ldots\right)=\frac{V_s}{\sqrt{2}}\frac{x\sqrt{2}}{x\sqrt{2}-1}$$

and

$$V_l = \frac{V}{\sqrt{2}}\left(1-\frac{1}{x\sqrt{2}}-\frac{1}{(x\sqrt{2})^2}-\frac{1}{(x\sqrt{2})^3}-\ldots\right)=\frac{V_s}{\sqrt{2}}\frac{x\sqrt{2}-2}{x\sqrt{2}-1}$$

and these lead to expressions identical to those already given in equations (1) and (2) for the power ratios in the accelerator and resistive load arms of the bridge when $n=1$. When the attenuation is 3 dB it takes five transits of the system to give 94% of the steady-state power build-up, and owing to the low group velocities of loaded waveguide this process may take an appreciable part of the radio-frequency pulse. This is a factor to be taken into consideration when estimating the overall efficiency of an accelerator using feedback. Inspection of the above series shows that as the loop attenuation increases more rapid convergence takes place, with consequent reduction in effective build-up time. Reduction in the value of n has a similar effect.

The chief point arising from this analysis is that the input power to an accelerator employing feedback will vary with the loop attenuation and hence with the beam loading. The behaviour of the bunching section under variable power input conditions will now be examined.

§ 3. The Bunching Section

The action of the bunching section of a linear accelerator has been described by Walkinshaw (1948). In this section a proportion of the electrons initially injected at low energies are brought to a longitudinally stable phase and travel along the remainder of the accelerator in 'bunches'. Theoretically, with a long enough section and a slow rate of rise of electric field strength with distance along the section, it is possible to trap nearly all the injected electrons into bunches. In practice, however, the limited radio-frequency power sources available and the desirability of economy in length in many applications demand a short section with high field strength in which the bunching action is rapid, but in which a high proportion of the injected electrons are 'lost'. The fact that all the injected electrons do not emerge in the high-energy beam does not matter provided that enough current is available from the injection gun to give the maximum output desired from the machine, and also, provided that electrons which are 'lost' have not first abstracted enough energy from the radio-frequency wave to cause appreciable attenuation. The latter is unlikely to be the case in a well-designed bunching section since electrons are presumably lost when they have fairly low axial velocities by deflection to the walls.

In practice a satisfactory design has been employed in which, theoretically, rather more than half the injected electrons are bunched at a phase angle of about 45° with respect to the zero of the radio-frequency wave. When the accelerated electrons attain an energy of the order of 1 Mev their velocity is nearly that of light, and little further bunching can take place. The bunches can then be made to ride near the crest of the radio-frequency wave to obtain maximum energy gain.

The design of a bunching section necessarily involves the assumption of a fixed distribution of phase velocity and electric field strength along its length. Hence it will only operate in accordance with the design for a fixed power input and beam current. If the input power varies from the design figure the action of the bunching section will be to try to maintain the electron energy at its output constant by altering the phase of the bunches. It is this characteristic of the bunching section which makes it different from the rest of the accelerator in which (as will be shown later) the energy gain falls as the power input is reduced or as the beam current is increased. When the power input to the bunching section is reduced the bunches will be formed nearer the crest of the wave and there will be a fall in the proportion of injected electrons which are trapped into the bunches. Eventually a power level is reached at which no effective bunching takes place and the behaviour of the section becomes unpredictable. In an accelerator employing feedback the input power falls continuously with increasing loading and so the process described above could effectively limit the useful output of the machine. Naturally, in practice, one designs the bunching section to be correct at the desired beam loading, and this involves a fairly precise knowledge of the properties of the waveguide employed and the attenuation of the various components of the accelerator-feedback loop.

§ 4. Waveguide Parameters

The important design parameters of the loaded waveguide used in linear accelerators are the series impedance and the attenuation per unit length, and in the case of the bunching section a knowledge of the variation of these properties with the phase velocity of the wave is also required. The group velocity is also of importance for estimating build-up time and frequency tolerance.

The series impedance $\xi = E^2/W$ relates the peak accelerating field E with the power flow W. A theoretical estimate of this parameter is difficult to obtain, and it is best to measure it using a perturbation technique similar to that described by Maier and Slater (1952). Waveguide of corrugation pitch 2 cm has been commonly used for linear accelerators in this country, and figure 3 shows a plot of measured values of series impedance against the radius of the iris hole for the particular

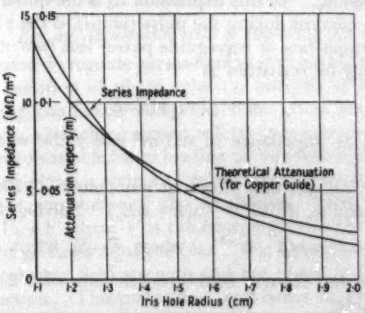


Figure 3. Series impedance and theoretical attenuation of corrugated waveguide of 2 cm pitch at a wavelength of 10 cm and when the phase velocity equals that of light.

case when the phase velocity of the propagated wave is equal to the velocity of light. Walkinshaw and Bell (1951, 1952) have shown that the field in the axial region of the guide can be expressed as an infinite series of space harmonics, and values for the amplitudes of these harmonics have been calculated for a particular case (Walkinshaw and Ross 1952), the computational effort being very great. Using these values, as described by Miller (1954), and integrating the Poynting vector over the guide cross section, a theoretical value for series impedance for this case has been worked out, and agreement with the experimental result was within 5%.

The attenuation of a waveguide is given, for small values, by $\alpha = w/2W$ nepers/m where w is the power loss per metre. It is related to the series impedance by the equation $\alpha = \xi/2\eta$; η is the shunt impedance and is more easily calculable since, as Walkinshaw and Ross (1952) have shown, it is not sensitive to the field matching approximations, and Walkinshaw's earlier theory (1948) may be used with reasonable accuracy. Theoretical values of the attenuation of copper guide at a wavelength of 10 cm and assuming a conductivity of $5{\cdot}8 \times 10^7$ mhos/metre are shown in figure 3, but in practice some allowance must be made for an increase in copper losses over the theoretical figure. This allowance depends on surface finish (Morgan 1949) and on constructional technique, and in practice an increase of from 25% to 50% has been found. The attenuation of corrugated waveguide may easily be measured (see, for instance, Mullett and Loach 1948).

§5. Beam Loading Theory

5.1. *For a Length of Guide with Constant Parameters*

We shall now consider the beam loading and energy gain of a length of waveguide with constant series impedance and attenuation and in which the electrons are injected with velocities approaching that of light. The phase of the electron bunch with respect to the radio-frequency wave is also assumed constant over the length.

Johnsen (1951) has employed a dimensionless beam-loading parameter which leads to some simplification of the equations. This parameter is the ratio of the initial rate of gain of beam power to the initial waveguide power loss per unit length, $m = (E_0 I \sin\phi)/w_0$. In this expression E_0 is the initial axial field strength (MV/m), I the mean current during the pulse (amps), ϕ the phase of the electron bunch and w_0 the initial rate of waveguide power loss (MW/m).

The equation may be rewritten as

$$m = (\xi I \sin\phi)/2\alpha E_0 \qquad \ldots\ldots(4)$$

with ξ denoting series impedance in MΩ/m² and α the waveguide attenuation in nepers/m.

The total power loss per unit length at any point is $-dW/dZ = w + EI\sin\phi$, and eliminating W and w, $dE/dZ = -\alpha(E + mE_0)$. Solving this equatiion,

$$E/E_0 = e^{-\alpha Z} - m(1 - e^{-\alpha Z}), \quad \text{where } E = E_0 \text{ when } Z = 0 \qquad \ldots\ldots(5)$$

and this shows how the electric field strength falls with distance Z along the accelerating waveguide.

If $E = E_1$ when $Z = L$ then $\ln(E_1/E)$ gives the attenuation of the length of accelerating waveguide under beam loaded conditions.

The energy gained by the electrons in traversing length L is given by

$$V = \int_0^L E \sin\phi \, dZ$$

and since we have assumed $\sin\phi$ to be constant then

$$\frac{V}{(E_0 \sin\phi)\, L} = \frac{1-e^{-\alpha L}}{\alpha L}(1+m) - m \qquad \ldots\ldots(6)$$

which shows that the energy gain decreases linearly with increasing loading for a fixed value of input power.

The power gain of the beam is

$$VI = \left[\frac{1-e^{-\alpha L}}{\alpha}(1+m) - mL\right] E_0 I \sin\phi$$

and substituting from equation (4) we get

$$\frac{VI}{W_0} = [(1-e^{-\alpha L})(1+m) - \alpha L m]\, 2m \qquad \ldots\ldots(7)$$

where W_0 is the power flow into the section in megawatts. This ratio may be thought of as the loading efficiency of the length of waveguide. It should be noted that a maximum is reached for a value of m such that

$$\frac{1}{m} = 2\left[\frac{\alpha L}{1-e^{-\alpha L}} - 1\right] \qquad \ldots\ldots(8)$$

and if this is substituted into equation (6) it is found that the energy gain at maximum efficiency is exactly half the no-load energy gain, that is, for a fixed input power. Obviously the smaller the value of αL the higher the maximum efficiency attainable, so that the best type of waveguide to employ is one that combines low attenuation with high series impedance, that is, it should have a high shunt impedance.

Substitution of the value of m given by equation (8) into the field equation (5), and making the approximation that $e^{-\alpha L} = 1 - \alpha L + \frac{1}{2}(\alpha L)^2$ it can be shown that at the loading giving maximum efficiency the radio-frequency field at the end of the accelerator has fallen to $(\frac{1}{2}\alpha L)^2 E_0$, that is, to a very low value. A further increase in loading must lead to the field falling to zero and building up in opposite phase, extracting power from the beam.

5.2. *For an Accelerator with a Bunching Section*

Since the corrugated waveguide parameters vary along the length of a bunching section due to the variation of phase velocity, analysis of its behaviour under variable beam-loading conditions along the lines given above is not possible. By its nature, however, a bunching section will try to produce electrons of constant energy when subject to variable beam loading or variable input power. We may therefore regard it simply as a source of relativistic electrons of fixed energy V_b for injection into the remainder of the accelerator. The final energy of the accelerator is then $V_b + V$ where V is the energy gain in the subsequent part of the accelerator and is given by equation (6). W_0, E_0 and ϕ are now respectively the power flux, the electric field strength and the phase at the beginning of the main accelerating section. This power flow, and hence the field strength, may be deduced, for any loading, from a knowledge of the attenuation α_b of the bunching section, making due allowance for the proportion of the total beam loading

occurring in this section. Then we have $W_0 = W_a \exp(-2\alpha_b)$, where W_a is the power input to the accelerator.

The final beam power may therefore be written as

$$\frac{(V+V_b)I}{W_a} = \frac{VI}{W_0}\frac{(1+V_b/V)}{\exp(2\alpha_b)} \qquad \ldots\ldots(9)$$

and VI/W_0 may be obtained from equation (7).

It is usually desirable to advance the phase of the bunches so that they ride nearer the crest of the radio-frequency wave when traversing the main accelerating section, and the corrugated waveguide may easily be designed to effect this change. The phase of formation of the bunches ϕ_b and the final phase ϕ will vary with any change of power level in the bunching section from the design figure. Such changes, whether due to alteration of input power or of beam loading, may be allowed for as an approximation by assuming $\sin\phi_b$ to be inversely proportional to the square root of the power flux averaged over the length of the bunching section, this being the condition for the maintenance of constant energy V_b. As pointed out in § 3 above, this procedure is only valid for power levels such that bunches are formed appreciably below the peak of the radio-frequency wave. This approximation involves some uncertainty as to the phase of the bunches, and it is intended to investigate the exact behaviour of bunching sections under variable input power conditions using an analogue computer (Crowley-Milling 1954).

5.3. *For an Accelerator employing Feedback*

When radio-frequency power feedback is used, the only modification to the theory outlined above is that we must determine the accelerator input power level as a function of beam loading. To calculate the power flowing in the bridge arms a knowledge of the attenuation of the components of the feedback loop is required. Let α_1 be attenuation of the rectangular waveguide and transformer at the accelerator input, and α_2 that of the output transformer and feedback waveguide, including the phase shifter. Let the steady state power levels in the bridge arms be W_a, W_f and W_s for the accelerator, feedback and input arms respectively, and W_b, W_0 and W_1 be the power levels in the corrugated waveguide at the beginning of the bunching section (if any) and at the beginning and end of the main accelerating section respectively. Then

$$W_a = W_b \exp(2\alpha_1) = W_0 \exp\{2(\alpha_1+\alpha_b)\},$$
$$W_f = W_1 \exp(-2\alpha_2),$$

and since from equation (6)

$$W_1 = W_0[e^{-\alpha L} - m(1-e^{-\alpha L})]^2,$$

then

$$\frac{W_f}{W_a} = [\exp\{-2(\alpha_1+\alpha_2+\alpha_b)\}]\,[e^{-\alpha L} - m(1-e^{-\alpha L})]^2.$$

Using the theory given in § 2, we can deduce that

$$\frac{W_a}{W_s} = \frac{1}{\{(n+1)^{1/2} - [\sqrt{n}\exp\{-(\alpha_1+\alpha_2+\alpha_b)\}]\,[e^{-\alpha L} - m(1-e^{-\alpha L})]\}^2} \qquad \ldots(10)$$

this being the equivalent form of equation (1).

Obviously, one may use the power relationships given here in conjunction with the equations above in order to determine beam power and energy gain in terms of the source power.

By way of illustration calculations have been made for a purely relativistic accelerator consisting of a 3 m length of corrugated waveguide of pitch 2 cm and iris hole radius 1·7 cm, the operating wavelength being 10 cm. The value of series impedance and attenuation are those derived from figure 3 with an increase of 37% in the attenuation value to give a practical figure. Since no assumptions have been made regarding source power or electron phase, we have calculated $V/W_s^{1/2}\sin\phi$, $(I\sin\phi)/W_s^{1/2}$ and the product VI/W_s. The values of $(I\sin\phi)/W_s^{1/2}$ are derived from values of m as defined by equation (4) which may be rewritten as

$$\frac{I\sin\phi}{W_0^{1/2}} = \frac{2\alpha}{\xi}m \quad \text{since} \quad E_0=(\xi W_0)^{1/2}.$$

When feedback is considered, of course, this must be multiplied by $(W_0/W_s)^{1/2}$ to give the beam current in terms of the source power.

The values of energy gain and beam power gain are shown plotted in figure 4

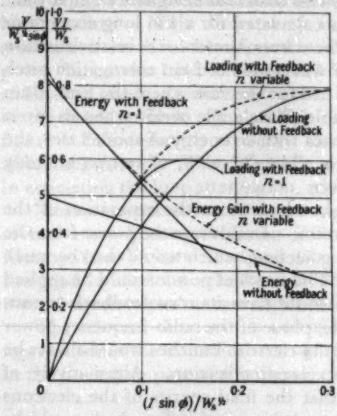


Figure 4. Electron energy gain and beam loading for a 3 m length of corrugated waveguide. Pitch 2 cm, iris hole radius 1·7 cm, wavelength 10 cm. V is in MeV, I in amperes and W_s is the source power in MW.

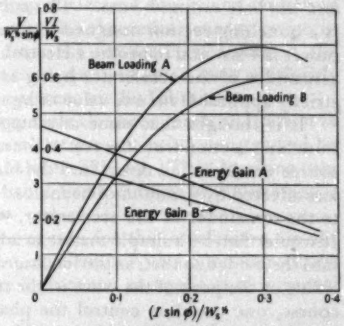


Figure 5. Electron energy gain and beam loading for a 1 m length of corrugated waveguide using feedback. A, no buncher; B, with buncher (for $W_s=2$ MW only). V is in MeV, I in amperes and W_s is the source power in MW.

against the beam current for two cases, one in which no feedback is used, the residual power being absorbed in a resistive load, and the second in which feedback is used with a unity ratio bridge. It will be noted that the use of feedback gives an appreciably higher energy gain at the lighter loadings, but that the energy gain falls more rapidly than it does without feedback, and eventually the curves cross at the point corresponding to a build-up ratio of unity (the 8 dB point in figure 2). At still higher loadings the bridge ratio is so far off its optimum value that the feedback is actually reducing the input power to the accelerator. Of course the employment of a bridge of correct n-value will always lead to some increase in efficiency at the corresponding loading provided that this is less than the value corresponding to maximum efficiency given by equation (8). To illustrate this point we have included in figure 4 curves showing the attainable

energy gain and beam power if a variable ratio bridge were used and always adjusted for optimum circulating power. These curves correspond to the one given in figure 2 for the maximum power ratio for any loop attenuation value.

For other fixed values of n lines may be drawn for energy gain which are tangential to the broken curve in figure 4 at the current loading value for which the bridge is correct and whose slope is steeper the higher the value of n.

The curves of figure 4 may be regarded as typical and show how the choice of feedback with bridge ratio suitable for a particular current loading affects the performance at other loadings. It should be emphasized that the results are only valid for a length of accelerator in which the electrons are injected in bunches at relativistic velocities. If a bunching section forms part of a short accelerator the energy gain will be less than in the relativistic case due to the formation of bunches at relatively low phase angle and also due to the reduction of series impedance occurring when phase velocities appreciably less than that of light are employed.

The effect is shown in figure 5, which is calculated for a 1 m long accelerator using feedback with a unity ratio bridge. The curves A refer to the relativistic case in which the whole length is constructed of waveguide of 2 cm corrugation pitch and iris hole radius 1·3 cm. The curves B are for the case where the first 30 cm is a bunching section designed to give a stable phase angle of 45° when its input power is 4 MW and to produce electron bunches with an energy of about 1 MeV, the remainder of the accelerator being as above. The curves B, of course, are only strictly applicable for one value of input power.

It is thought that some advantage might result from the separation of the bunching section from the rest of the accelerator. In this case the power from the source would be fed direct into the bunching section, which would then be much less affected by variation of beam loading. The residual power would be applied to the remainder of the accelerator, which would have its own feedback circuit. It would then be a simple matter to adjust the phase of the radio-frequency power into the bridge so that, as the loading varies, the electron bunches would always be riding at the peak of the wave in the main accelerating section. Alternatively, of course, one could so control the phasing that the final energy of the electrons remains constant with variation of loading, and this is a feature which would be desirable in some applications.

§ 6. Comparison of Theoretical with Measured Performance of Two Accelerators

The theory given in this paper has been used to calculate the expected performance of two linear accelerators designed for X-ray therapy equipment. In such accelerators it is found that the X-ray output reaches a maximum at medium loading, this being due to the high rate of variation of X-ray conversion in the target with electron energy (the exponent is about 2·7), so that high values of beam power output are not the primary concern.

The first equipment on which a comparison has been made is the Metropolitan-Vickers 8 MeV linear accelerator (Miller 1953 a, b, 1954). This consists of 3 metres of corrugated waveguide with overall feedback. The bunching section occupied the first 40 cm and was designed to have an output energy of 1·3 MeV, and the remainder of the accelerator had a corrugation pitch of 2 cm and an iris hole radius of about 1·68 cm. The input power from the source assumed in the design was about 1·86 MW peak. Figure 6 shows the calculated energy as a function of beam

current together with the measured points for the energy at the peak of the spectrum. Widths of the spectra between half-value points were of the order of 0·3 MeV. It is estimated that the uncertainties in the calculation and in the measurement involve a possible error of ± 0·3 MeV. Within these limits the agree-

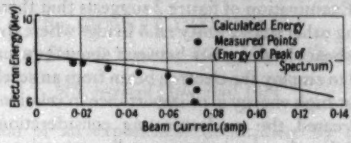


Figure 6. Comparison of theoretical and experimental performance of an 8 MeV linear accelerator.

ment at light and medium loadings is good, but a serious discrepancy occurs at high loadings. This discrepancy is, however, attributable to an anomalous attenuation which has been found to occur in the accelerator over a certain range of power flux, and it has been shown (Miller and Saxon 1953, Miller 1954) that this anomalous effect can account for the sharp drop of electron energy and limitation of the beam current. It should be noted that the theoretical energy curve in figure 6 differs slightly at heavy loadings from the similar curve given in the references above. This difference was due to an approximation used previously in estimating the loop attenuation, which the present method eliminates.

Measurements of beam current and energy have also been made on the first of several 4 MeV linear accelerators ('Orthotrons') at present being constructed by Metropolitan-Vickers Electrical Co. Ltd. (Miller 1953 b). This machine was designed to accelerate electrons to 4 MeV with a beam current of 200 mA and a radio-frequency peak power input of 2·0 MW.

The relevant data on the corrugated waveguide for this accelerator are the figures used in the numerical example for a 1 m accelerator with a bunching section in the preceding part of this paper. Figure 7 gives the comparison of expected

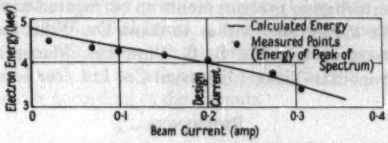


Figure 7. Comparison of theoretical and experimental performance of a 4 MeV linear accelerator.

performance with the experimentally determined points for the energy at the peak of the spectrum at various beam loadings, and a good agreement is observable over the range of loading used. It is hoped later to extend the energy measurements to higher beam currents. It should perhaps be mentioned that the spectrum widths were of the order of 0·15 to 0·2 MeV between half intensity points in the majority of the measurements, though somewhat wider spectra are obtained at the lighter loadings where the difference between the experimental points and the curve in figure 7 is greatest. Anomalous attenuation effects have apparently been avoided, presumably due to the employment of higher field strengths than in the 8 MeV machine.

§ 7. Conclusions

In so far as it is possible to generalize it may be stated that the employment of feedback with bridge ratios of the order of unity has definite advantages in applications where high-energy gain in a short length is the primary requirement as in x-ray equipments. Examination of figure 2 suggests that there would seem to be little reason for making other than a unity ratio bridge where the loop attentuation, including the desired beam loading, lies between about 2 dB and 5 dB.

When it is desired to employ the electron beam from an accelerator, for instance for sterilization, whilst high energy is of importance in relation to the thickness of the material to be treated, the predominating consideration may well be the economic one of deriving the maximum overall conversion of radio-frequency power into beam power. Assuming that, in practice, the theoretical maximum efficiency could be reached, the use of feedback would, of course, be pointless; but, as demonstrated in figure 4, at loadings less than the value giving optimum efficiency there will always be some advantage in using feedback with low bridge ratios.

From the comparison given in § 6 it may also be concluded that it would appear possible to design linear accelerators with a reasonable degree of precision once the relevant corrugated waveguide properties are known. This remark must be qualified, however, in that the anomalous attenuation which is observed in the 8 MeV linear accelerator, and which may well be present in other similar machines, especially where lower power levels are employed, is far from being fully understood and some research is required on the best means of avoiding this effect.

Acknowledgments

The author has attempted to show his indebtedness to other workers in this field by his choice of references, but he would like particularly to acknowledge valuable discussions of the problems involved in linear accelerator design with his colleagues Messrs. C. W. Miller and M. C. Crowley-Milling and with Mr. W. Walkinshaw of the Atomic Energy Research Establishment, who have made helpful suggestions regarding this paper. Thanks are also due to Dr. A. W. Aikin who carried out the perturbation measurements on corrugated waveguide mentioned in the paper. The author also wishes to thank Dr. Willis Jackson, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

References

Crowley-Milling, M. C., 1954, *J. Sci. Instrum.*, **31**, 100.
Johnsen, K., 1951, *Proc. Phys. Soc.* B, **64**, 1062.
Maier, L. C., and Slater, J. C., 1952, *J. Appl. Phys.*, **23**, 68.
Miller, C. W., 1953 a, *Nature, Lond.*, **171**, 297; 1953 b, *Metrop. Vick. Gaz.*, December; 1954, *Proc. Instn Elect. Engrs*, Pt. I, **101**, 207.
Miller, C. W., and Saxon, G., 1953, *Nature, Lond.*, **172**, 463.
Morgan, S. P., 1949, *J. Appl. Phys.*, **20**, 352.
Mullett, L. B., and Loach, B. G., 1948, *Proc. Phys. Soc.* B, **61**, 271; 1949, *A.E.R.E. Report* EL/R. 391.
Shersby-Harvie, R. B. R.-, 1948, *Proc. Phys. Soc.* B, **61**, 255,
Shersby-Harvie, R. B. R.-, and Mullett, L. B., 1949, *Proc. Phys. Soc.* B, **62**, 270.
Walkinshaw, W., 1948, *Proc. Phys. Soc.* B, **61**, 246.
Walkinshaw, W., and Bell, J. S., 1951, *A.E.R.E. Report* G/R. 675; 1952, *A.E.R.E. Report* T/R. 864.
Walkinshaw, W., and Ross, M., 1952, *A.E.R.E. Memorandum* T/M. 55.

A Monochromatically Ionized Layer in a Non-Uniformly Recombinant Atmosphere; with Applications to the D and E Ionospheric Regions.†

By SYDNEY CHAPMAN
Geophysical Institute, University of Alaska

MS. received 13*th April* 1954

Abstract. Ionization of an exponential atmosphere by monochromatic solar radiation is considered when the recombinance α is not independent of height but varies as $\alpha_0+\alpha_1 e^{-bh}$, where α_0, α_1 and b are constants. A *recombinance datum* level is defined, namely the level where α_0 and $\alpha_1 e^{-bh}$ are equal: heights are measured from this level, in scale-height units. The level of the *absorption peak* being z_χ (when the sun's zenith distance is χ, or z_0 when $\chi=0$), the level z_m of the electron peak and the height distribution of the electron density n_e are considered, for different values of z_0 and of $c(=b/H)$, particularly $c=1$, 2, 3. When $c>0$ the electron peak is always above the absorption peak, and for $c\geqslant 2$ it is always above the recombinance datum level: the electron peak for $c=1$, when the absorption peak is below the recombinance datum level, is about half way between the two. The decrease of n_e (from its maximum value n_{em}) on the underside (or incline) of the electron layer can be much less steep than for a Chapman layer, if the absorption peak is below the recombinance datum level.

The results for the model atmospheres considered are tentatively discussed with reference to the E and D ionospheric regions, but their potential value may be realized only when better data for the D region become available.

§ 1. Introduction; Non-Uniform Recombinance

This Note supplements my previous papers (Chapman 1931 a, b, 1939) on model atmospheric ionized layers. It differs from them as regards α, the recombination coefficient, which for brevity I shall term the *recombinance*. Whereas previously I took α to be independent of the height h, here I suppose that it increases downwards according to the formula

$$\alpha=\alpha_0+\alpha_1 e^{-bh} \qquad \ldots\ldots(1)$$

where α_0, α_1 and b are positive constants. Nicolet (1951) has considered another type of varying recombinance in a more general type of atmosphere.

The height h' at which the variable and constant parts of α are equal is given by $h'=(1/b)\ln(\alpha_1/\alpha_0)$; this height h' will be called the *recombinance datum* level for an atmosphere of this type. It is convenient to rewrite equation (1) in the form

$$\alpha=\alpha_0[1+\exp\{-b(h-h')\}] \qquad \ldots\ldots(2)$$

As previously, I consider an *exponential* atmosphere, uniformly mixed, with constant scale height H, so that the number density of each constituent varies as $\exp(-h/H)$.

It is convenient to introduce the new height-measure z defined by

$$z=(h-h')/H \qquad \ldots\ldots(3)$$

† Completed while the author was visiting professor at New York University.

reckoned in scale-height units from the recombinance datum level. Hence (2) will be replaced by

$$\alpha = \alpha_0(1 + e^{-cz}), \qquad \ldots\ldots(4)$$

where

$$c = b/H. \qquad \ldots\ldots(5)$$

According to (4), the case $c = 0$ corresponds to $\alpha = 2\alpha_0$ for all values of z, but here the case $c = 0$ will be taken to imply $\alpha_1 = 0$ and $\alpha = \alpha_0$ in (1).

A parallel beam of monochromatic solar radiation, which may be absorbed by more than one constituent, ionizes one of them, whose number density n will be expressed by

$$n = n' e^{-z} \qquad \ldots\ldots(6)$$

where n' denotes its number density at the datum level h'.

The atomic or molecular absorption coefficient of this ionized constituent will be denoted by k; hence kn, at the level z, will be its (ordinary) absorption coefficient, here called the *absorptance*.

Any other absorbing constituents, 1, 2, ..., not ionized by this radiation, will contribute $k_1 n_1$, $k_2 n_2$, ... to the total absorptance A, and the general molecular scattering A_s will also contribute. Absorption of scattered radiation, and refraction, will be neglected. Thus the total absorptance A is given by

$$A = \sum kn + A_s = A_0 e^{-z}; \qquad \ldots\ldots(7)$$

all the terms in A vary as e^{-z}, and A_0 denotes the total absorptance at the recombinance datum level.

If the zenith distance of the sun is χ, and the atmosphere is plane-stratified (the curvature of the level layers being neglected), there is a maximum rate of absorption at the height h_χ given by $h_\chi = h_0 + H \ln \sec \chi$; here h_0 denotes the value of h_χ for $\chi = 0$ (that is, for the sun overhead); it is given by

$$h_0 = H \ln (A_0/H). \qquad \ldots\ldots(8)$$

This level h_χ will be called *the absorption peak*, by analogy with my previous proposals as to upper atmospheric nomenclature (Chapman 1950 a, b); the level h_0 may be called the zenith absorption peak; let

$$z_\chi = (h_\chi - h')/H, \qquad z_0 = (h_0 - h')/H. \qquad \ldots\ldots(9)$$

The rate of ionization q ($\mathrm{cm^{-3}\,sec^{-1}}$) will be given, in terms of its maximum value q_χ (or q_0 when $\chi = 0$), corresponding to the absorption peak, by

$$q = q_\chi \exp[1 - (z - z_\chi) - \exp\{-(z - z_\chi)\} \sec \chi]; \qquad \ldots\ldots(10)$$

and q_χ is given by

$$q_\chi = \beta I_\infty \cos \chi / He = q_0 \cos \chi \qquad (e = 2{\cdot}718\ldots) \qquad \ldots\ldots(11)$$

where I_∞ denotes the number flux of the photons in the solar stream, per $\mathrm{cm^2}$ per sec, outside the atmosphere, and β is a fraction denoting the proportion of the absorbed photons that produce an electron–ion pair from the ionized constituent. (If each photon absorbed by this constituent produces such a pair,

$$\beta = k n'/A_0). \qquad \ldots\ldots(12)$$

The total number Q of electrons formed per $\mathrm{cm^2}$ per sec by the absorbing beam will be given by

$$Q = \beta I_\infty = eHq_\chi \sec \chi = eHq_0. \qquad (e = 2{\cdot}718\ldots) \qquad \ldots\ldots(13)$$

If desired, the curvature of the level layers of the atmosphere can be taken into account by replacing $\sec\chi$ and $\cos\chi$ in the above expressions by the function denoted by $\mathrm{Ch}(x, \chi)$ and its reciprocal (Chapman 1931 b, 1953, Wilkes 1954).

§ 2. The Electronic Equilibrium

The equation of change of the electron density n_e is taken to be

$$dn_e/dt = q - \alpha n_e^2 . \qquad \ldots\ldots(14)$$

This may be expressed (Chapman 1931 a) in terms of one parameter (σ_0), proportional to $(\alpha q_0)^{-1/2}$; in the D and E regions of the ionosphere, to which the present analysis may be applicable, this parameter is sufficiently small for the value of n_e to approximate closely throughout the day hours, except near dawn and sunset, to the equilibrium value corresponding to the then existing rate of production

$$n_e = (q/\alpha)^{1/2}. \qquad \ldots\ldots(15)$$

The height distribution of this equilibrium value will here be considered, particularly for the case $\chi = 0$ (overhead sun); the modification for oblique sunlight is easily made. Writing

$$n_{e0} = (q_0/\alpha_0)^{1/2}, \qquad n_{e\chi} = (q_\chi/\alpha_0)^{1/2} \qquad \ldots\ldots(16)$$

and using equations (4), (10), n_e for $\chi = 0$ can be expressed in the form

$$\frac{n_e}{n_{e0}} = \frac{\exp\left[\frac{1}{2}\{1-(z-z_0)-\exp\left[-(z-z_0)\right]\}\right]}{(1+e^{-cz})^{1/2}} ; \qquad \ldots\ldots(17)$$

for zenith angle χ the ratio $n_e/n_{e\chi}$ is given by the same expression with z_0 replaced by z_χ. Note that for $z = 0$ (17) is independent of c, and depends only on z_0.

§ 3. The Electron Peak and Electron Distribution

The level z_m (or the height h_m, equal to $h' + Hz_m$) at which the electron density n_e has its maximum value n_{em} will be called the *electron peak*. It is given by $d\ln n_e = 0$ or, using (15), by $d\ln q = d\ln\alpha$, which leads to the equation

$$\frac{e^z\{1-(c-1)e^{-cz}\}}{1+e^{-cz}} = \exp z_0 \qquad \ldots\ldots(18)$$

whose solution is z_m.

If $c = 0$, so that the recombinance is independent of the height, as supposed in my earlier papers, the solution is $z_m = z_0$, that is, the electron peak coincides with the absorption peak; and n_{em} is then n_{e0}.

If c is positive, as here supposed, the left-hand side of (18) is less than e^z, and therefore $z_m > z_0$: that is, *the electron peak always lies above the absorption peak*.

When the absorption peak is several scale heights above the recombinance datum level, so that cz_0 (and therefore also cz_m) is large (e.g. 3 or more), an approximation to (18) is

$$z_m - z_0 = c\exp(-cz_0) \quad (cz_0 > 3). \qquad \ldots\ldots(19)^*$$

In this case n_{em} differs very little from n_{e0}, and the electron layer differs little from a Chapman layer (which corresponds to $c = 0$ or α uniform); the difference between the two lies in the slightly greater height of the electron peak (by the above small amount $c\exp(-cz_0)$, and a slightly smaller electron density in the 'tail' of the underside (or 'incline') of the layer.

* The equations whose numbers have asterisks are *approximate*, not exact.

At the other extreme, when the absorption peak lies several scale-heights *below* the recombinance datum (so that z_0 is negative) $\exp z_0$ is very small, and the approximate behaviour of z_m depends on whether c is greater or less than 1. If $0<c<1$, the left-hand side of (18) is positive for all values of z, and (18) can be satisfied only by a large negative value of z; its approximate value is given by

$$\left.\begin{aligned} z_m - z_0 &= -\ln(1-c) \\ \text{and} \quad n_{em}/n_{e0} &= (1-c)^{1/2}\, e^{c/2} \end{aligned}\right\} \quad (0<c<1;\ cz_0<-3). \qquad \ldots\ldots(20)^*, (21)^*$$

When $c=1$,

$$z_m = \tfrac{1}{2}z_0, \quad n_{em}/n_{e0} = \exp\{\tfrac{1}{2}(1+z_0)\} = 1{\cdot}649\exp(\tfrac{1}{2}z_0) \quad (c=1;\ z_0<-6). \qquad \ldots\ldots(22)^*$$

When $c>1$, and z_0 is large and negative, (18) is satisfied by a value of z that makes the numerator of the fraction on the left-hand side small; and the approximation to z_m is

$$z_m = \frac{1}{c}\ln(c-1) + (\exp z_0)\left(\frac{1}{c-1}\right)^{(c+1)/c} \quad (c>1;\ -cz_0 \text{ large}) \qquad \ldots\ldots(23)^*$$

$$n_{em}/n_{e0} = \{c^{-1}(c-1)^{1-1/c}\exp(1+z_0)\}^{1/2} \quad (c>1;\ -cz_0 \text{ large}). \qquad \ldots\ldots(24)^*$$

When $c=2$ these approximate formulae become

$$z_m = -\tfrac{1}{2}\exp z_0, \quad (c=2;\ -z_0 \text{ large}) \qquad \ldots\ldots(25)^*$$

$$n_{em}/n_{e0} = \{\tfrac{1}{2}\exp(1+z_0)\}^{1/2} = 1{\cdot}166\exp(\tfrac{1}{2}z_0) \quad (c=2;\ -z_0 \text{ large}). \qquad \ldots\ldots(26)^*$$

In this case (18) can be expressed exactly (for any value of z_0) in the special forms

$$\exp\{-(z-z_0)\} = \tanh z, \quad z = z_0 + \ln\coth z \quad (c=2), \qquad \ldots\ldots(27)$$

which enable z_m to be determined easily, using tables of exponential and hyperbolic functions. When $c=3$ the formulae become

$$z_m = \tfrac{1}{3}\ln 2 + 2^{-4/3}\exp z_0 = 0{\cdot}231 + 0{\cdot}397\exp z_0 \qquad \ldots\ldots(28)^*$$

$$n_{em}/n_{e0} = 2^{1/3}\{\tfrac{1}{3}\exp(1+z_0)\}^{1/2} = 1{\cdot}199\exp(\tfrac{1}{2}z_0). \qquad \ldots\ldots(29)^*$$

When $z_0=0$ and $c=1$, z_m is $\sinh^{-1}\frac{1}{2}$, namely 0·481; when $z_0=0$ and $c=2$, z_m is the solution of the equation $z=\ln\coth z$, namely 0·6095; when $z_0=0$ and $c=3$, z_m is the solution of the equation $\sinh 2z = \cosh z + \frac{1}{2}e^{-z}$, namely 0·585. Thus z_m for $z_0=0$ does not vary monotonically with c. The corresponding values of n_{em}/n_{e0} in the three cases are respectively 0·75, 0·81 and 0·86.

At z_0 (the absorption peak) n_e/n_{e0} is $(1+e^{-cz})^{-1/2}$, or, when z_0 is large and negative, approximately $\exp(-\frac{1}{2}cz_0)$; this approximation is valid to within about one part in 1000 for $-cz_0>3$.

These results are illustrated in figures 1–4. Figure 1 shows how z_m varies with z_0 for four values of c, namely 0, 1, 2, 3; for $c=0$, $z_m=z_0$. For the other three values of c the graphs of z_m asymptotically approach the line $z_m=z_0$ for large positive values of z_0; for $c=1$ the graph asymptotically approaches the line $z_m=\frac{1}{2}z_0$ for large negative values of z_0; for $c=2$ and $c=3$, z_m approaches a constant limit for large negative values of z_0, this limit being zero for $c=2$, and $\frac{1}{3}\ln 2$ or 0·231 for $c=3$. Thus for $c=2$ the electron peak for radiation whose absorption peak is well below the recombinance datum level tends (from above) to identity with the absorption peak; but for $c=3$ it tends to a limit about $\frac{1}{4}H$ above this datum level.

Figure 2 shows more clearly, and in more detail, the variation of z_m with z_0 for these values of c; for $z_0 \geqslant 0$ it gives the graphs of $z_m - z_0$; for $z_0 \leqslant 0$ it gives, for $c=1$, the graph of $z_m - \frac{1}{2}z_0$, and for $c=2$, 3 it gives the graphs of z_m itself.

Figure 3 shows graphs of n_{em}/n_{e0} as a function of z_0 for $c=1$, 2, 3; for no value of z_0 is the difference between the three graphs great, and for $z_0 < -1$ the difference between the graphs for $c=2$, 3 is very small. Also for large positive values of z_0 the difference is in all cases small.

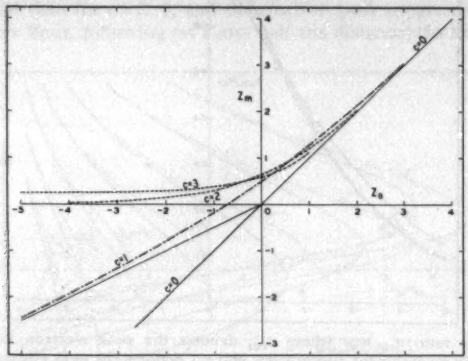


Figure 1. Levels z_m of the electron peak as a function of the level z_0 of the absorption peak for a monochromatically ionized layer of an exponential atmosphere in which the recombinance α is given by $\alpha_0(1+e^{-cz})$ for $c=0$, 1, 2, 3. The levels z, z_0, z_m are measured in scale-height units from the recombinance datum level, at which the constant and variable parts of α have the same value.

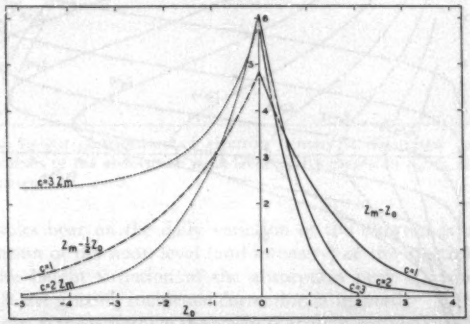


Figure 2. The levels z_m of the electron peak as a function of the level z_0 of the absorption peak, as given in figure 1, are here shown in more detail, by the graph, for z_0 positive, of $z_m - z_0$, and for z_0 negative by the graphs of z_m, for $c=2$, $c=3$, and of $z_m - \frac{1}{2}z_0$, for $c=1$.

Figure 4 shows curves of n_e/n_{e0} as a function of z for $c=2$ for a series of values of z_0, namely -5 to 4. For a given beam intensity, the deeper the penetration the smaller is the electron peak intensity n_{em}, and the lower the electron peak;

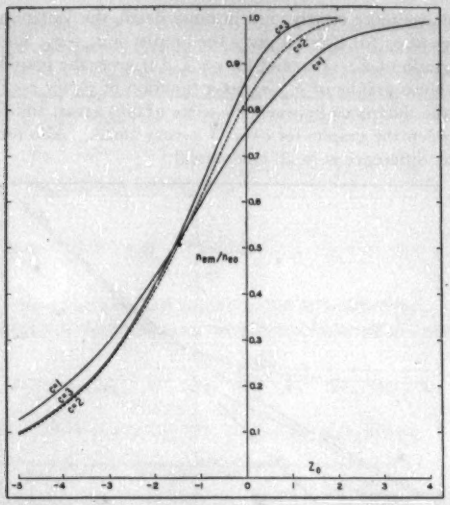


Figure 3. The ratio n_{em}/n_{e0}, where n_{em} denotes the peak electron density in a non-uniformly recombinant atmosphere, and n_{e0} denotes the peak electron density in the corresponding case for height-independent recombinance, is shown as a function of the level of the absorption peak, z_0, for three values of the recombinance parameter c, namely 1, 2, 3.

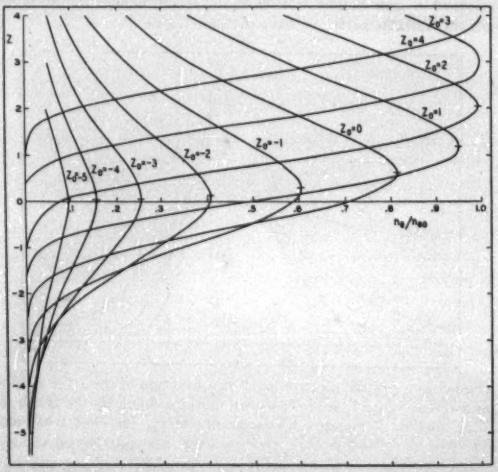


Figure 4. The height distribution of the electron density n_e, as a fraction of the maximum value n_{e0} for height-independent recombinance, is shown as a function of level z for several values of the level z_0 of the absorption peak, for a non-uniformly recombinant atmosphere for which $c=2$.

but this does not descend below the recombinance datum level. The cut-off of the ionized layer on its underside or incline is steeper than for constant recombinance ($c=0$), when $z_0 \geqslant 1$, though the difference rapidly dies away as z_0 increases; for $z_0 \leqslant 0$ the layer incline is *less* steep than for $c=0$, and the difference increases with $-z_0$, that is, the lower the absorption peak.

Figure 5 shows the curves of n_e/n_{e0} as a function of z for $c=1$ (full lines) and $c=3$ (broken lines); for $c=1$ and z_0 negative the curvature of the graphs is more gradual than for $c=2$, 3, and the electron peak progressively descends, without a lower limit, following (at about half the distance) the absorption peak.

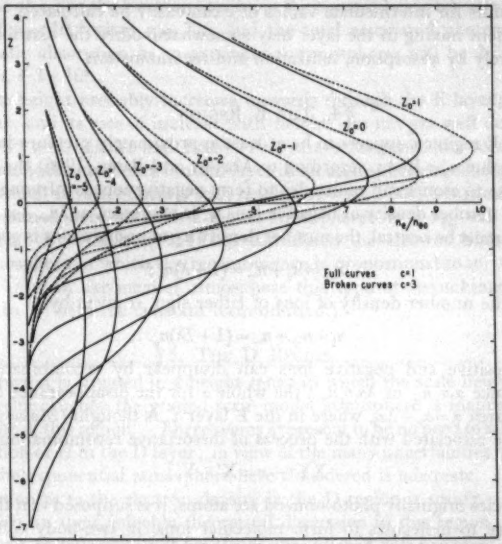


Figure 5. The height distribution of electron density is illustrated as in figure 4 for several values of the absorption peak level z_0, by graphs of n_e/n_{e0} as a function of z, for the cases $c=1$, $c=3$.

These results bear on the daily variation of the electron layer, and on the seasonal variation of the noon level (and intensity) of the electron peak; this is because of the height variation of the absorption peak z_χ, which is equal to $z_0+\ln\sec\chi$ (if the grazing incidence correction is ignored).

For example, at the equator at the equinox at noon, $\sec\chi=1$, and the absorption peak is at some level z_0; it is higher by one scale height at the time when $\sec\chi=e$, or $\chi=68{\cdot}4°$. At 45° latitude at the equinox at noon $\sec\chi=1{\cdot}414$, and the noon absorption peak is at $z_\chi=z_0+\frac{1}{2}\ln 2=z_0+0{\cdot}346$; the absorption peak is one scale height higher than this when χ is about 75°. The corresponding changes of level of the *electron* peaks will be less than one unit of z (or H) by amounts which depend on the level z_0 and on c; for $c=2$ or 3 the change will be small if the zenith absorption peak lies well below the recombinance datum.

The peak value of q is proportional to $\cos\chi$, but the peak value of the electron density is not proportional to $(\cos\chi)^{1/2}$ because n_{em}/n_{e0} is greater the higher the absorption peak. Suppose, for example, that $z_0=-2$ at the equator at the equinox at noon; later in the day, when χ has increased to 68·4° and z_χ is -1, $n_{e\chi}$ is then $e^{-1/2}n_{e0}$; but for $c=2$, n_{em}/n_{e0} is increased from about 0·41 for $z_0=-2$ to about 0·61 at $z_0=-1$; hence n_{em} is reduced, in the interval from noon to the epoch when $\chi=68{\cdot}4°$, only in the ratio $(0{\cdot}61/0{\cdot}41\,e^{1/2})$ or about 0·90. Only if z_0 is 1 or more does n_{em} vary during the sunlit hours approximately as $(\cos\chi)^{1/2}$.

When more accurate data for the D region are provided by the radio physicists, such results as these deserve detailed comparison with them. If found desirable, similar results for intermediate values of c can easily be calculated.

Turbulent mixing in the layer may somewhat modify the distribution determined purely by absorption, ionization and recombination.

§ 4. The E Region

In the E region α appears to be at least approximately pressure-independent, with the value $\alpha_0=10^{-8}$; according to Massey and Bates (1946) the attachment of electrons to atoms and molecules to form negative ions is infrequent there, so that if the number density of negative ions n_- is expressed as λn_e, λ is less than 1. As the gas must be neutral, the number density n_+ of positive ions is given by

$$n_+=n_e+n_-=(1+\lambda)n_e;$$

hence n_i, the number density of ions of either sign, is given by

$$n_i=n_++n_-=(1+2\lambda)n_e.$$

The positive and negative ions can disappear by recombination, with a recombinance $\alpha_i n_- n_+$ or $\lambda\alpha_i n_e n_+$; the whole α for the disappearance of electrons can be written $\alpha=\alpha_e+\lambda\alpha_i$, where in the E layer α_e is thought (Massey and Bates 1946) to be associated with the process of dissociative recombination

$$XY^+ + e = X' + Y';$$

if the particles originally photo-ionized are atoms, it is supposed that they transfer their charge to molecules to form molecular ions, in two-body collisions. In the E layer it is supposed that α_e is the main term in α. If photodetachment nearly balances detachment of electrons during the day, λ will be approximately proportional to n.

According to Mitra (1952, p. 260) the E layer extends down to 88 km. Just above this level, at 90 km, λ is estimated as 2 by Bates and Seaton (1950), who also quote an estimate 3×10^{-8} for α at 90 km by W. R. Piggott. If these estimates are valid, the recombinance datum level h' defined in § 1 is above 90 km and lies in the E layer. Taking α_0 as 10^{-8}, and α at 90 km as 3×10^{-8}, the second term in α in equations (1) and (4) must there be twice the first; hence h' must lie at a level $(H/c)\ln 2$ above 90 km; taking H as 7 km, and $c=1$, 2 or 3, the height h' is found to be 94·9, 92·4 or 91·6 km. As the E electron peak is about 4 scale heights above this level h', z_0 also must be about 4. Hence the E layer, if it were formed by monochromatic radiation, would have its electron peak almost exactly at the level of the absorption peak, and would be indistinguishable from a 'Chapman' layer. At 90 km, approximately 4 scale heights below the E peak, the electron density of such a layer should be less than the peak value by the

factor $\exp(1+4-e^4)$ or about 10^{-21}; actually Bates and Seaton quote an estimate (privately communicated) by W. R. Piggott that n_e at 90 km is $1{\cdot}5\times10^4$, or 10^{-1} times the peak value. This is incompatible with monochromatic absorption, even allowing for the downward increase of α, which should further decrease n_e at 90 km.

If molecular oxygen is the constituent whose photo-ionization maintains the E region, it may be ionized mainly by x-rays (Byram, Chubb and Friedman 1953), but partly by the band from 1029 to 910 Å. The less absorbed part of this band may increase the electron density below the E peak, possibly down to 90 km: for a model case of band absorption see Chapman (1939).

The value of n_{em} in the E region at noon at the equator being taken as $1{\cdot}5\times10^5$, the value of $\alpha_0=10^{-8}$ gives $q_0=225$, and the total associated photon flux, for monochromatic absorption in an exponential atmosphere, will be $2{\cdot}718q_0H$ or, for $H=7$ km, $4{\cdot}3\times10^8$.

The scale height probably increases upwards through the E layer; its value at the E peak, and its rate of increase with height, are not yet well determined. Menzel (1947) and Nicolet and Bossy (1949) have discussed the mathematical theory of a monochromatically ionized layer in an atmosphere of constant recombinance, in which H varies linearly with height: $H=H_0+b'h$; I suggest the name *graded-scale* atmosphere for such a model atmosphere; in it n_{em} will vary with χ as $(\cos\chi)^{(1+b')/2}$. Nicolet (1951) has considered the electron distribution in such an atmosphere in which the recombinance is proportional to n^c, where c is a constant. (In an exponential atmosphere this type of recombinance would correspond to (4) with the constant term omitted.)

§ 5. The D Region

The D region is situated in a height range in which the scale height H has a minimum value; hence the graded-scale model atmosphere is inappropriate in the discussion of this region. There seems at present to be no need to take account of the variation of H in the D layer; in view of the many uncertainties concerning this layer, the exponential atmosphere here considered is adequate.

Information as to the electron density in the D region is scanty. Bates and Seaton (1950), in their valuable theoretical discussion of this region, quoted an estimate that n_e is 250 cm^{-3} at 75 km (from unpublished data privately communicated by K. Weekes). Gardner and Pawsey (1953), using new sensitive methods at a station near Sydney, Australia, found a distinct stratum with an electron peak of about 250 cm^{-3} at about 72 km, and a region of greater electron density at about 90 km that extends up to the E region; this region descends as the sun rises, and ascends at night, but does not disappear. The latter is the region that will here be discussed as the D region; the lower stratum is variable from day to day.

Gardner and Pawsey give tentative graphs of n_e in the region below 90 km for three epochs in the day–noon, near sunset, and after sunset: the last of these suggests a peak at about 88 km at that time. The noon peak, if any, is not indicated, but may be at about 80 km; the peak value of n_e is not given. Gnanalingham and Weekes (1952) find, by a less sensitive method, that n_e in summer at Cambridge, England, cannot exceed about $1{\cdot}3\times10^4$ below 100 km.

From the estimate $n_e=250$ at 75 km, and the attainment *near noon* of the maximum value of n_e at this level, Bates and Seaton (1950), using my theoretical

curves for n_e as a function of time, infer that at this level q (noon) and α must be at least 6×10^{-2} and 10^{-6} respectively; 10^{-6} is 500 times the value of the estimated variable part (2×10^{-8}) of α at 90 km, 2·2 H higher; the ratio 500 is expressible as $e^{2 \cdot 2c}$, which gives $c = 2 \cdot 3$; but little reliance can be placed on this estimate of c.

If the D region is formed by absorption of monochromatic radiation, and if $c \geqslant 2$, it appears (§ 3) that the electron peak cannot descend below the noon absorption peak for overhead sun; at other solar zenith angles χ the absorption peak will be higher by the amount $H \ln \sec \chi$, and the electron peak will also be higher.

The estimates quoted for n_e and α at 90 and 75 km are not all for the same place, and are perhaps rather uncertain. The Sydney results perhaps suggest an absorption peak for overhead sun at about 80 km, which (assuming monochromatic ionization) would imply that the recombinance datum level h' is not above 80 km, unless $c < 2$. It is perhaps more likely that h' is at about 90 km, and that the 'after sunset' n_e peak shown by Gardner and Pawsey (in their figure 18) is doubtful: and that c is 2 or perhaps somewhat greater. If so, the D peak would at all times be above about 90 km, but owing to the tenfold smaller value of n_e there, the peak might not be observable against the background of the overlying E region.

This conclusion, however, may not be applicable to the solar-controlled ionized layer, with its peak at about 75 km, found by Gardner and Pawsey, even if this layer is due to monochromatic photo-ionization, because the recombinance for this layer may follow a different law, and have a different recombinance datum level, from those appropriate to the main D region. Such differences might exist, for example, if this region is ionized by the absorption by atomic sodium of radiation in the band 2415–2200 Å. Sodium ions cannot transfer their charge to any molecule in the atmosphere, so that the recombinance cannot be dissociative; possibly it may be by one of the following processes, after the electrons have become attached to molecular oxygen:

$$Na^+ + O_2^- = NaO + O, \quad \text{or} \quad Na^+ + O_2^- = Na + O_2;$$

the former is perhaps the more likely. Bates and Seaton (1950), who discuss the sodium hypothesis of the origin of the D region, advanced by Jouaust and Vassy (1941) and Vassy and Vassy (1942), conclude that sodium could not give adequate ionization except at the extreme base of the D layer; perhaps sodium is responsible for the low peak found by Gardner and Pawsey, though the rapid fluctuations of the reflection they got from this level does not well fit the sodium hypothesis, unless the solar radiation absorbed by the sodium fluctuates greatly. Bates and Seaton remark that the sodium-absorbed radiation is not appreciably attenuated by such absorption; if so, the variation of q with height will not be according to (10), but will be determined by the distribution of atomic sodium; in this case the considerations in §§ 2, 3 will not necessarily be applicable to the layer.

Bates and Seaton, after discussing other hypotheses as to the origin of the D layer, favour the one advanced by Nicolet (1945), that the constituent there photo-ionized is nitric oxide, NO (ionization potential 9·2 ev, and that the ionizing radiation is the Lα (Lyman alpha) line of hydrogen, centred at 1216 Å; this can penetrate below the E region owing to a partial atmospheric 'window' between the O_2 Schumann–Runge absorption bands and the strong bands of O_2 and N_2

beyond 1000Å; the amount of NO present in the D region is not known, but Nicolet suggests that it is of uniform concentration, owing to mixing overcoming diffusive separation; so that its absorption (taking H to be constant) will have the form (10).

Recently (Byram, Chubb, Friedman and Lichtman 1952) Lα solar radiation has been detected by rockets in the D region; the inferred intensity outside the atmosphere is of the order 10^{10} to 3×10^{10} photons $cm^{-2} sec^{-1}$ for quiet sun. During solar flares the D region is much more strongly ionized, presumably by a large enhancement of the breadth and total intensity of Lα associated with the ground-observed enhancement of the Hα line. This favours the hypothesis that the D region is due to substantially monochromatic radiation; according to Sun and Weissler (1952) and Watanabe, Inn and Zelikoff (1952) the absorption coefficient of NO at Lyman alpha is $67{\cdot}5\ cm^{-1}$ (at STP) and varies only between this value and 55 over the range 1200–1226Å. Ditchburn and Heddle (1935) found a rather smaller value, $47\ cm^{-1}$, giving rather deeper penetration, and agreeing better with the value inferred from the rocket observations. Over the same range the absorption coefficient for O_2 varies between $0{\cdot}3\ cm^{-1}$ (at Lyman alpha) to 500 at 1205Å and does not exceed 13 between 1216 and 1230Å; thus, if Lyman alpha is broadened during a solar flare, only the short-wave wing will be specially affected by O_2.

The NO^+ ions cannot transfer their charge to any other molecule, but may recombine with an electron with dissociation; they may also recombine with oxygen molecular ions in a two-body process. It may be that in the D region c is 1 instead of 2.

References

Bates, D. R., and Seaton, M. J., 1950, *Proc. Phys. Soc.* B, **63**, 129.

Byram, E. T., Chubb, T., and Friedman, H., 1953, *Phys. Rev.*, **93**, 1066.

Byram, E. T., Chubb, T., Friedman, H., and Lichtman, S. W., 1952, *J. Opt. Soc. Amer.*, **42**, 876.

Chapman, S., 1931 a, *Proc. Phys. Soc.*, **43**, 26; 1931 b, *Ibid.*, **43**, 483; 1939, *Ibid.*, **51**, 93; 1950 a, *J. Atmos., Terr. Phys.*, **1**, 121; 1950 b, *J. Geophys. Res.*, **55**, 395; 1953, *Proc. Phys. Soc.* B, **66**, 710.

Ditchburn, R. W., and Heddle, D. W. O., 1953, *Proc. Roy. Soc.* A, **220**, 61.

Gardner, F. F., and Pawsey, J. L., 1953, *J. Atmos. Terr. Phys.*, **3**, 321.

Gnanalingham, S., and Weekes, K., 1952, *Nature, Lond.*, **170**, 113.

Jouaust, R., and Vassy, E., 1941, *C. R. Acad. Sci., Paris*, **213**, 139.

Massey, H. S. W., and Bates, D. R., 1946, *Proc. Roy. Soc.* A, **187**, 211.

Menzel, D. H., 1947, *Rec. Un. Radio-Sci. Int.*, **6**, 132.

Mitra, S. K., 1952, *The Upper Atmosphere* (Calcutta: Asiatic Society).

Nicolet, M., 1945, *Mem. Inst. Mét. Belg.*, **19**, 124; 1951, *J. Atmos. Terr. Phys.*, **1**, 141.

Nicolet, M., and Bossy, L., 1949, *Ann. de Géophys.*, **5**, 275.

Vassy, A., and Vassy, E., 1942, *Cah. Phys.*, **9**, 28.

Watanabe, K., Inn, E. C. Y., and Zelikoff, M., 1952, *J. Chem. Phys.*, **20**, 1969.

Wilkes, M. V., 1954, *Proc. Phys. Soc.* B, **67**, 304.

RESEARCH NOTES

The Temperature Variation of Susceptibility of Tantalum

By F. E. HOARE, J. S. KOUVELITES†, J. C. MATTHEWS and J. PRESTON

The Physics Laboratories, University of Leeds

MS. received 31*st May* 1954

In a previous communication (Hoare and Walling 1951) the results were given of a determination of the absolute susceptibility of tantalum at 20°C. Small specimens turned from the rod then used have since been employed to obtain room temperature susceptibilities of a number of metals and alloys. Kriessman (1953) has published an account of measurements on the high temperature variation of susceptibility of several metals including tantalum. The mass susceptibility at 25°C is given as $0{\cdot}827 \times 10^{-6}$, which is to be compared with our value of $0{\cdot}8490 \times 10^{-6}$ at 20°C. The difference between these two results is much larger than the probable error of the latter value. As both specimens were of spectrographically standardized material supplied by Johnson, Matthey and Co. Ltd., it is unlikely that the divergence in values arises from significant differences in the impurity content. Kriessman gives no details of how his absolute value was obtained and no assessment of its accuracy.

Since the temperature variation of susceptibility is very small, the change in value to be expected over the temperature range 20–293°K is much smaller than that of the majority of metals. The photographic method initially adopted for measurement has been described previously (Hoare and Matthews 1952). After several runs the photographic method was abandoned in favour of a servo-mechanism operated by a split photocell which enables measurements to be obtained more rapidly. The accuracy obtainable by the photographic and servo-mechanism methods is comparable, as has been shown by comparative tests on several metals. Since the absolute accuracy of both methods of measurement is independent of the susceptibility being measured, the percentage error in the result for a metal such as tantalum is higher than for metals with a large susceptibility.

Over six hundred measurements of the susceptibility of tantalum have been made between 20 and 293°K. It was thought at first that the best representation of those observations would be obtained by a straight line, and straight lines were fitted by least squares to each run. Some runs did not cover the whole temperature range and it was found that, on fitting straight lines to the results of these, the temperature coefficient was positive or negative, depending upon whether it was the lower or upper half of the relevant temperature range which was being considered. This was so consistent that the observations on runs covering the complete temperature range were divided into groups corresponding to the two halves of the temperature range and separately analysed. The results

† Now at Division of Applied Science, Harvard University, Cambridge, Mass., U.S.A.

were consistent with the observations for partial runs and indicated the possibility of a maximum in the susceptibility as has been established in the case of palladium.

Consequently all the results were taken together and a quadratic form fitted by least squares. The final expression for the susceptibility, in the temperature range from 20 to 293°K, relative to that at 293°K, is

$$\chi \text{ rel} = 1{\cdot}001 + 6{\cdot}457 \times 10^{-5}\, T - 2{\cdot}315 \times 10^{-7}\, T^2.$$

Out of more than six hundred results 97·5% lie within 1% of the curve and 81% within 0·5%. The standard deviation is 0·004.

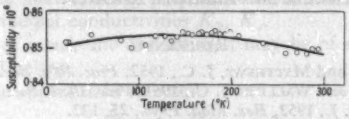


Figure 1. The susceptibility of tantalum. The circles give mean values of every twenty consecutive experimental results.

It is hardly possible to reproduce in graphical form, with a suitable scale, all the experimental results. To give some idea of the scatter, however, figure 1 has been prepared, in which each point plotted corresponds to the mean value of a group of twenty observations, except at the extremities of the range where more observations are included in each group. The full curve gives the absolute susceptibility calculated from the equation above and our value at 293°K. The distribution of points and their divergence from the curve give an idea of the frequency and accuracy of observations in different temperature ranges.

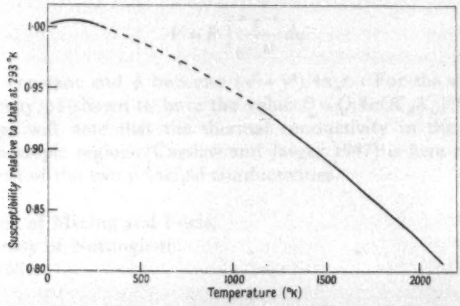


Figure 2. The relative susceptibility of tantalum. The lower temperature portion of the curve represents the present work, and the upper temperature portion that of Kriessman, both relative to 293°K. For the dotted portion no published data can be found.

Owing to the differences in absolute values the present results have been compared with those of Kriessman in terms of relative susceptibility. Through the courtesy of Dr. Kriessman we have available a large-scale reproduction of his results. From this the relative values have been taken off over the temperature

range 293 to 2000°K and are shown plotted in figure 2 together with the curve computed from the above equation. It will be seen that the continuity of the total curve is satisfactory.

In conclusion it should be emphasized that the evidence for the existence of a maximum in the susceptibility of tantalum is *very* much weaker than in the case of palladium. The scatter of the experimental points obscures the maximum which only appears after the mathematical analysis.

ACKNOWLEDGMENT

J.C.M. and J.P. wish to express their gratitude for maintenance grants from the Department of Scientific and Industrial Research.

REFERENCES

HOARE, F. E., and MATTHEWS, J. C., 1952, *Proc. Roy. Soc.* A, **212,** 137.
HOARE, F. E., and WALLING, J. C., 1951, *Proc. Phys. Soc.* B, **64,** 337.
KRIESSMAN, C. J., 1953, *Rev. Mod. Phys.*, **25,** 122.

LETTERS TO THE EDITOR

Transient Heat Flow in Anisotropic Strata

The writer has recently solved the particular problem of transient heat flow from an infinite line source of strength Q heat units per unit length per unit time embedded in an infinite region exhibiting two-dimensional anisotropy, the direction of the line source being perpendicular to the plane containing the directions of the principal conductivities K_x, K_y.

As the method of attack and the solution may be of more general interest they are reproduced below:

Starting with the appropriate equation

$$K_x \frac{\partial^2 v}{\partial x^2} + K_y \frac{\partial^2 v}{\partial y^2} = \rho C \frac{\partial v}{\partial t}$$

where ρ and C are respectively the density and specific heat of the region (assumed constant) and v is the temperature at point P (x, y) and at time t, we make the transformation $\bar{y} = y(K_x/K_y)^{1/2}$ and obtain the equation

$$\alpha_x \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial \bar{y}^2} \right) = \frac{\partial v}{\partial t}$$

where $\alpha_x = K_x/\rho C$ is the thermal diffusivity in the direction of x.

We can now make use of the standard solution in *cylindrical* coordinates for an infinite line source in isotropic regions, viz.

$$V = B \int_{\phi}^{\infty} \frac{e^{-u}}{u} du$$

where B is a constant and ϕ becomes $(x^2 + \bar{y}^2)/4\alpha_x t$. For the anisotropic case the constant may be shown to have the value $B = Q/4\pi(K_x K_y)^{1/2}$.

The reader will note that the thermal conductivity in the corresponding solution for isotropic regions (Carslaw and Jaeger 1947) is here replaced by the geometric mean of the two principal conductivities.

Department of Mining and Fuels, D. R. Scott.
University of Nottingham.
7th July 1954.

Carslaw, H. S., and Jaeger, J. C., 1947, *Conduction of Heat in Solids* (Oxford: Clarendon Press), p. 221.

REVIEWS OF BOOKS

Nature and Structure of Collagen, edited by J. T. RANDALL. Pp. ix+269. (London: Butterworths Scientific Publications, 1953). 42*s*.

Few materials in the animal body are of greater importance alike for the understanding of disease processes as well as normal ones than the connective tissue substance called Collagen. It is also of some industrial significance in relation to leather, glue and photographic emulsions. It comes as no surprise therefore to find that a symposium on Collagen held at the Biophyiscs Department, King's College, London, attracted a large audience of biologists, chemists and physicists, and that 26 papers were presented. The meeting represented an informal activity of the Colloid and Biophysics sub-committee of the Faraday Society, but the proceedings did not, unfortunately, rank for publication in the well-known 'General Discussion' reports of that Society. This difficulty has been surmounted by publication in attractive book form by Butterworths, and a great many who were unable to attend the conference will be grateful for this.

However, the reader must not expect too much. It must be borne in mind that the exact investigation of this tantalizing substance is among the most difficult of present-day biophysical and biochemical tasks. Collagen is difficult to obtain in solution and it is doubtful whether the material which dissolves represents the whole or only the outside of the natural fibre. Also the soluble portion has different properties, depending on whether it is extracted in acid, neutral or alkaline solutions. As was stressed at the conference, there is not yet even a satisfactory terminology with which to define the different fractions, and consequently the properties reported by one observer need not necessarily confirm those by another—a perplexing difficulty which is particularly evident in papers dealing with Collagen precursors.

No soluble fraction of Collagen shows evidence of molecular uniformity comparable to that characterizing the so-called homogeneous proteins, e.g. albumins or haemoglobins, and the ultra-centrifuge, electrophoresis and light-scattering studies reported here appear not to alter this situation in any important respect. On the other hand, in its morphological aspects Collagen—or rather tendon—gives results of great beauty, as is evidenced by the striking electron-micrographs and x-ray diffraction pictures to be found in this volume. The proceedings led up to a final and all too brief discussion of the atomic arrangement or 'crystallography' of tendon. It is evident that the atomic spirals of Pauling, Corey, Bear and others dominate the picture, but with significant qualifications, e.g. that repeating 3-amino acid residues in the spiral need not be strictly equivalent, and that the amino acid composition of the various helices may not be the same. The stumbling biologist is still left with the suspicion that Collagen may not be a reproducible chemical entity after all, but may in its molecular structure always reflect in some measure the chemical life history of the animal from which it came. It seems that physical and chemical techniques which have enriched this field of study so much in the past decade must yet be further refined before this point can be finally decided.

A. S. MCFARLANE.

Twinning and Diffusionless Transformations in Metals, by E. O. HALL. Pp. 181. (London: Butterworths Scientific Publications, 1954). 30*s*.

This is not a good book, yet this reviewer will keep it at hand on his shelf, and probably consult it fairly often. It is valuable because it contains much the largest available collection of facts and references about its two subjects. The general plan of the book is good and the theoretical treatment occasionally felicitous, but a host of irritating errors hides these merits from view. One feels that if the author had submitted his manuscript to a few colleagues before publication the overall impression made by the book could have been very different.

The worst chapter is the first, a chapter of elementary crystallography abounding with minor errors and half truths, and serving onlv to illustrate the rapidity with which a science may decay if it ceases to develop. At the first page it is apparent that the author does not really know what a lattice is. How widespread is his illusion that the smallest repeated unit in the face-centred cubic lattice contains four atoms? And what does one who holds this belief take to be the basis of Bravais's classification of lattices? The table of crystal systems, said to be reproduced by permission from Barrett, needlessly varies his wording and falsifies one definition by omitting a 'not'. Newly drawn figures illustrating crystal structures are generally badly drawn, e.g. figure 8 (most unfairly attributed to A. Taylor, who draws it right) and figure 7*b*, which is gravely misleading into the bargain. Figure 56 (tin) shows the wrong composition plane, apart from other drawing errors, while the projections of the cadmium and magnesium structures in figures 51 and 52 are thoroughly and incomprehensibly wrong—a serious matter when the atomic movements in twinning are the subject of discussion.

On p. 47 the author claims to correct an error in Schmid and Boas—he actually produces the same result as they do, but in a more cumbersome form, though the treatment as far as the last three lines is a neater one. At the same time he reproduces (as figure 36) figure 70 from their book, still with its original error (an interchange of IV and VI in either figure or legend). Another amendment of Schmid and Boas given on p. 82 appears to be a proper one.

On p. 50, under 'Determination of the Indices of Twinning', the author says that when relative orientations have been determined "inspection . . . then indicates that the poles of the twin can be derived from the poles of the matrix by reflection in a single plane, which must therefore be the composition plane K_1"; and immediately follows this by an example (α-iron) for which there are four such reflection planes, three of them being shown in his diagram.

As the reviewer has already indicated, the book is not really as bad as that: but it is definitely a book for the research worker, who will verify statements when they are of importance to him. It should be kept out of the hands of students: above all, let no one attempt to start learning crystallography from that first chapter.

F. C. FRANK.

High Altitude Rocket Research, by HOMER E. NEWELL, Jr. Pp. xiv+298. (New York: Academic Press; London: Academic Books, 1953.) $7.50.

It is unusual to find that the central theme of a research report is not so much a specific field of study, or even a special technique of investigation, but rather the means of transport of the investigating instruments. The present book is

a review of attempts at direct measurement by rocket-borne instruments (mainly from the Proving Ground in New Mexico) of as many as possible of the characteristics of the atmosphere above 30 km, including pressure, density, temperature, composition, ion- and current-density, ultra-violet insolation and cosmic radiation. Although these features are not of course unrelated, the main unity lies in the limitations of all rocket-borne investigations—the extremely high cost and the short period for which instruments may be exposed. There is evidence that (as we should expect) the atmosphere is as variable at high as at low levels so that the most should be made of the results. The author rightly pays special attention to the errors to which this work is susceptible, and from his account it is possible to make a fair estimate of the reliability of the inferences. Some of the data derived (e.g. ultra-violet spectrograms) are otherwise unobtainable. Other results may be compared with those obtained by indirect methods, but the author deliberately avoids any detailed comparison. Apart from the resulting one-sidedness the book is quite a good introduction to the physics of the upper atmosphere since each section contains enough 'background' to make it immediately intelligible to the general reader.

E. T. EADY.

Mathematics in Action, by O. G. SUTTON. Pp. viii+226. (London: Bell, 1954.) 16*s*.

Dr. Sutton describes the nature of applied mathematics and gives examples of its use. The book is written for the layman but the reviewer feels that a knowledge of mathematics up to Intermediate standard is necessary for comprehension. Apart from a chapter on statistics, the examples are mainly taken from the admittedly broad field of applied mechanics. The reviewer personally found the book well written and interesting.

D. R. BLAND.

Electronics : Physical Principles and Applications, by ARTHUR O. WILLIAMS, Jr. Pp. v+306. (New York: D. Van Nostrand, 1953.) 34*s*.

The choice of topics covered in this volume is unusual for a book on electronics. As the sub-title indicates, physical principles and applications are considered, while the actual valve circuits are only given two chapters. The author starts from the fundamental concepts in electricity such as the definitions of current and potential difference. Circuit analysis is considered next, and properties of R–C, R–L, and oscillatory circuits are discussed. After a chapter on amplifiers and one on oscillators, modulation and detection are discussed under the name of 'processing of information'. Then follows a treatment of electromagnetic wave propagation. In the last hundred pages applications are discussed, in particular radiometry, spectroscopy and ion optics. Short introductions are given to optics, as well as atomic and molecular structure.

This extensive syllabus is covered in about 300 pages and naturally the treatment is not very deep. The material is presented from the conceptual point of view and only a beginner's knowledge of calculus is required. The arguments are rendered plausible rather than proved. The chapters on applications are written in a lively manner and could almost be read independently of the rest of the book. Unfortunately the chapters on principles suffer from many undesirable features.

The comparison of the definitions of potential difference and electromotive force (p. 34) shows that they are stated in a fallacious manner: "The difference of potential between two points is defined as the work required to carry a positive charge from one point to the other divided by the magnitude of the charge." The electromotive force is defined as "the work required to carry a positive charge completely round the circuit divided by the magnitude of the charge". In view of Kirchhoff's laws these definitions imply that no e.m.f. can exist in a closed circuit, whether the circuit contains a battery or not.

Kirchhoff's 'Rule 2' is stated as follows (p. 37): "Taken in a single direction around any loop, the sum of potential drops equals the sum of potential rises." Yet the author claims on page 45 that there exist conditions under which this rule breaks down, in particular switching transients and circuits whose dimensions are comparable with or greater than the relevant wavelengths. But it is well known that the electrostatic potential is a univalued function of space under all circumstances. The author claims that the second law of Kirchhoff tacitly assumes an infinite velocity of propagation.

Throughout the book voltages in circuits are not properly specified: they appear as definite expressions in equations but are not assigned a direction in the circuit diagrams. This leads to arbitrariness in the mathematical formulation and is particularly undesirable in the treatment of rotating vectors; any one of these vectors could in fact be drawn in the opposite direction to the one given. Furthermore no meaning can be attached to the concept of a leading or a lagging phase unless the directions of both currents and voltages are specified. On page 75 the rotating vectors are introduced, but no statement appears as to the axis on to which the vectors are to be projected. It is not even stated whether they should all be projected on the same axis.

On page 60 the half-life of a radioactive material is incorrectly defined; on page 80 a statement appears that a capacity does not on the average dissipate any power; this seems to suggest that dissipation of power is a reversible process. Feedback is introduced on page 96; the principal diagram is lamentably wrong, as the input and the 'fed-back' voltages are connected in parallel. Space does not permit further enumeration of misleading statements.

It is the reviewer's opinion that the selection of topics in this volume is suitable for a book to be recommended as a qualitative introduction to electronics for science students who do not specialize in physics. A. FOLKIERSKI.

CONTENTS OF SECTION A

PAGE

Dr. Ernest Bauer and Dr. Ta-You Wu. The Cooling of a Gas by Radiation . 741

Dr. P. J. Grant, Dr. F. C. Flack, Dr. J. G. Rutherglen and Mr. W. M. Deuchars. Gamma Radiation from the Reaction $^{11}B(p, \gamma)^{12}C$. . . 751

Mr. D. M. Brink. Short Range Forces and Nuclear Energy Levels in the Neighbourhood of ^{208}Pb 757

Dr. K. M. Gatha, Mr. G. Z. Shah and Mr. N. J. Patel. Approximate Nuclear Density Distributions in Light Elements 773

Dr. M. E. Pillow. Intensity Distribution among Nitrogen Bands in the Auroral Spectrum 780

Mr. W. G. Henry. The Self-Consistent Field for Au^+ 789

Prof. M. J. S. Dewar and Prof. H. C. Longuet-Higgins. The Electronic Spectra of Aromatic Molecules: I—Benzenoid Hydrocarbons 795

Prof. D. R. Bates and Dr. B. L. Moiseiwitsch. Inelastic Heavy Particle Collisions Involving the Crossing of Potential Energy Curves: I—Charge transfer from H atoms to Be^{2+}, Si^{2+} and Mg^{2+} ions 805

Mr. J. Yoccoz. Coulomb Effects in Stripping Reactions 813

Mr. C. R. Stanley. A New Method for the Production of Active Nitrogen and its Application to the Study of Collision Effects in the Nitrogen Molecular Spectrum 821

Dr. M. J. Buckingham and Dr. M. R. Schafroth. The Specific Heat of Metals at Low Temperatures 828

Dr. H. M. Rosenberg. The Thermal Conductivity of Germanium and Silicon at Low Temperatures 837

Mr. A. Ward. The Emission of Light in the Passage of Alpha Particles through Gases 841

Research Notes:

Dr. M. E. Pillow. Intensities in Band Systems of O_2: $^3\Sigma_u^+$–$^1\Sigma_g^+$ (Broida–Gaydon System) and $^1\Sigma_g^+$–$^3\Sigma_g^-$ (Atmospheric System) . . 847

Mr. P. C. Price. The Yield Curve of the $^7Li(p, \gamma)^8Be$ Reaction . . 849

Contents of Section B 852

The Lattice Spacings of Binary Tin-Rich Alloys

By J. A. LEE and G. V. RAYNOR
Department of Physical Metallurgy, The University, Edgbaston, Birmingham, 15

MS. received 8th June 1954

Abstract. The variation with composition of the lattice spacings of the solid solutions in tin of antimony, bismuth, lead, indium, cadmium, zinc and mercury has been determined. Antimony and bismuth decrease the axial ratio of white tin, while the other elements investigated increase the axial ratio. This suggests that the Brillouin zone for white tin is overlapped across planes of energy discontinuity which lie parallel to the tetragonal axis; the most probable form of this zone is discussed. The suggested Brillouin zone is bounded by planes of the (220) and (211) families, of which only the former are overlapped. The axial ratio variations may then be interpreted, and a possible explanation given for anomalies in the lattice spacing–composition curves for tin–indium and tin–cadmium alloys, which are associated with the development of vacant lattice sites.

§ 1. Introduction

Research on alloys of copper and silver (Hume-Rothery, Lewin and Reynolds 1936, Owen and Roberts 1939), magnesium (Raynor 1940, 1942, Hume-Rothery and Raynor 1940) and aluminium (Axon and Hume-Rothery 1948) has shown that the mean lattice distortion produced by a solute in substitutional solid solution may be regarded as the resultant effect of two factors:

(i) an 'atomic' factor: a solute of smaller atomic diameter than the solvent tends to contract the lattice, while a solute of larger atomic diameter tends to expand the lattice;

(ii) a 'valency' factor: the lattice tends to be expanded by a solute of higher valency than the solvent, and contracted by a solute of lower valency.

In alloys based on copper and silver, which are monovalent and correspond with a half-filled Brillouin zone, regularities are observed between the lattice distortion produced by equiatomic percentages of various solutes, and the atomic sizes and valencies of the solutes (Hume-Rothery, Lewin and Reynolds 1936, Owen and Roberts 1939, Raynor 1949). For magnesium as solvent the same general principles apply but regularity is lost. In the first Brillouin zone for magnesium overlap of electrons occurs only at right angles to the hexagonal axis; it is the variation of this overlap, or the initiation of new overlaps parallel to the hexagonal axis, which exerts the dominant influence on the lattice spacings of magnesium alloys (Hume-Rothery and Raynor 1940). Overlap of electrons also occurs from the first Brillouin zone for trivalent aluminium, and the lattice spacing relationships in aluminium-rich alloys are complex. Influences due to

relative ionic sizes, and relative volumes available per valency electron for the solvent and solute metals have, however, been traced (Axon and Hume-Rothery 1948).

It may be suggested that lattice spacing measurements may be used to obtain information about the probable electronic constitutions of solvent metals for which mathematical analyses prove difficult. To examine the probable Brillouin zone structure for metallic tin, therefore, the lattice spacings of a representative selection of tin-rich alloys have been measured. The results are described below.

§ 2. Materials and Methods

The following substitutional solid solutions have been studied: Sn–Sb, Sn–Bi, Sn–Pb, Sn–In, Sn–Cd, Sn–Zn, and Sn–Hg. In addition, the thermal expansion of white tin was measured between 9·5° and 182°c.

Alloys were prepared in small quantities, using the method outlined below, from the following materials: (i) Chempur tin, 99·99% pure; (ii) antimony and bismuth, not less than 99·95% pure; (iii) lead, mercury and cadmium, of purity not less than 99·99%; (iv) indium and zinc of spectrographically pure grade.

The technique for alloy preparation was similar to that of Owen (1943). Degreased clippings of the component metals were weighed into thin-walled Pyrex tubes each furnished with a bulb sufficiently large to accommodate the material. Each tube was sealed after having been flushed with argon, and evacuated. It was then attached to the rotating arm of a mixing device, and held for thirty minutes in the centre of a furnace controlled at 5°c above the melting point of the higher melting component. The liquid metal was tipped into the tubular portion of the preparation capsule, withdrawn from the furnace and quenched into cold water to ensure a fine-grained structure. The Pyrex tubes did not shatter; homogenization treatments were therefore carried out in the same tube, using controlled resistance furnaces. The annealing period was decided only after preliminary tests on each alloy system had indicated that uniformity of composition was obtainable in a period shorter than that subsequently employed.

Negligible weight losses occurred when alloys were melted and annealed under the conditions outlined. Intended compositions were therefore accurately attained, as confirmed by the chemical analysis of all alloys.

All annealed alloys were examined micrographically before the preparation of x-ray specimens. Filings were taken from both ends of the ingot, and from the faces exposed by halving the remainder of the ingot longitudinally; specimens from these samples were exposed separately to x-rays as a further check for homogeneity, and any alloys showing inconsistent results were rejected. The filings were mixed with gum-tragacanth, and the moistened mixture extruded through an orifice of diameter 0·3 to 0·5 mm. All x-ray specimens used for work at room temperature were thus obtained in a reproducible manner. Filings for high-temperature diffraction experiments were enclosed in evacuated thin-walled silica capillaries.

In the course of the work, numerous tests were made on samples of filings from the same ingot which had received different annealing treatments. No significant lattice spacing differences were observed, indicating no appreciable composition change on annealing. It is thus justifiable to take the compositions of the filings as accurately represented by those of the parent ingots.

Specimens were exposed to CuKα radiation, using Debye–Scherrer cameras of the van Arkel type, or a Unicam high-temperature powder diffraction camera calibrated against the lattice spacings of pure silver. The temperatures of the exposures were controlled to within $\pm\frac{1}{2}$°C. At high temperatures, the techniques and precautions discussed by Berry, Henry and Raynor (1951) were adopted. All lattice spacing values were expressed in kx units at 25°C, and the following constants were assumed:

(i) Wavelength of copper radiation: $K\alpha_1 = 1{\cdot}537\,395$ kx units,
$K\alpha_2 = 1{\cdot}541\,232$ kx units.

(ii) Lattice spacing of silver: $a_0 = 4{\cdot}077\,87$ kx at 25°C.

(iii) Linear coefficient of thermal expansion of silver: $\beta = 18{\cdot}8 \times 10^{-6}$.

(iv) Linear coefficient of thermal expansion of tin: $\beta_\perp = 16{\cdot}7 \times 10^{-6}$,
$\beta_\parallel = 36{\cdot}4 \times 10^{-6}$.

The lattice spacings were not corrected for refractivity.

§ 3. Experimental Results

3.1. *Lattice Spacings and Thermal Expansion of White Tin*

Three specimens of tin were exposed at room temperature. The mean of the three measurements gave $a = 5{\cdot}8199 \pm 0{\cdot}0002$ kx, $c = 3{\cdot}1749 \pm 0{\cdot}0002$ kx, $c/a = 0{\cdot}54552$ (25°C). For the same temperature, Ievens, Straumanis and Karlsons (1938) have reported $a = 5{\cdot}81970 \pm 0{\cdot}00002$ kx, $c = 3{\cdot}17488 \pm 0{\cdot}00005$ kx. The lattice spacings at elevated temperatures are plotted in figure 1; the mean coefficients of linear expansion over the range investigated are $\beta_\perp = 16{\cdot}7 \times 10^{-6}$, $\beta_\parallel = 36{\cdot}4 \times 10^{-6}$. These values agree substantially with macroscopic measurements (Erfluig 1939), and were used in the present work to reduce measured spacings to 25°C.

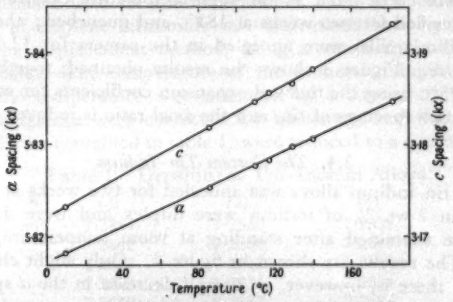


Figure 1. Lattice spacings of tin at elevated temperatures.

3.2. *The Systems Tin–Antimony and Tin–Bismuth*

Alloys were annealed for seven days at 220° or 225°C. Filings were re-annealed for strain-relief (two hours). In all cases three specimens were examined, and the lattice spacings measured agreed within the limits of experimental error ($\pm 0{\cdot}0002$ kx).

The lattice spacings are plotted against composition in figure 2. Both *c* and *a* spacings lie on straight lines. Although *c* for tin–antimony alloys varies little, *a* expands, and the axial ratio decreases continuously. For tin–bismuth alloys both *c* and *a* expand considerably. The axial ratio is again decreased.

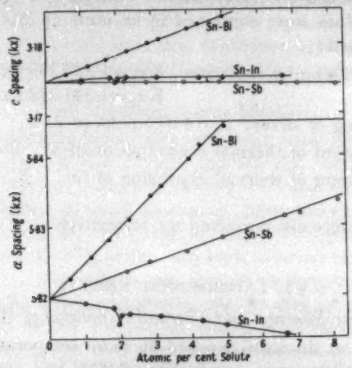


Figure 2. Lattice spacings in the systems tin–antimony, tin–bismuth and tin–indium (25°c). All compositions established by analysis.

3.3. *The System Tin–Lead*

The solid solubility of lead in tin is only 2·6 wt. % at the eutectic temperature of 183°c (Raynor 1947); the room-temperature solubility is very limited. Since the homogeneous high-temperature structure could not be retained on quenching, experiments were carried out in the high-temperature camera at $185 \pm 0{\cdot}5$°c. Ingots were annealed for two weeks at 183°c and quenched; the filings, sealed in evacuated silica quills, were annealed in the camera for 12 hours at 185°c prior to exposure. Figure 3 shows the results obtained, together with values calculated for 25°c using the thermal expansion coefficients for white tin. Lead expands the lattice spacings of tin, and the axial ratio is reduced slightly.

3.4. *The System Tin–Indium*

A series of tin–indium alloys was annealed for two weeks at 175°c. Alloys with more than 7 wt. % of indium were duplex and were discarded. The remainder were examined after standing at room temperature, *in vacuo*, for six months. The results are shown in figure 2. Only slight changes occur in the *c* spacing; there is, however, an overall decrease in the *a* spacing, and *c*/*a* increases. Between 1·8 and 2·4 atomic % of indium an anomaly occurs in the lattice spacing–composition relationships. Both spacings are affected, but *a* is reduced to a greater extent than *c*. Analogous lattice spacing anomalies have previously been observed, and have been attributed to the presence of vacant lattice sites. To investigate whether vacancies occurred in the present instance, further alloys were prepared for density measurements, using a slight modification of the method of Egerton and Lee (1923), for which an accuracy of one part in ten thousand for a 1-gramme specimen has been claimed.

Alloys were prepared as for the x-ray work, and vacuum cast. Although specimens appeared free from porosity, they were, as a precaution, reduced in thickness by 50% by squeezing between two polished, lightly lubricated aluminium plates. After cleaning and washing in alcohol, the specimens were annealed *in vacuo* for one week at 175°C. The densities were determined after four months

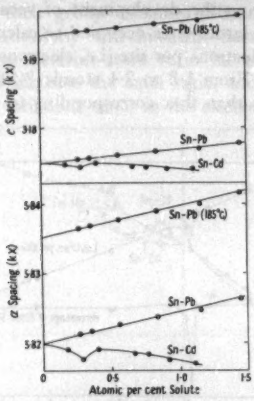


Figure 3. Lattice spacings in the systems tin–lead (185° and 25°C), and tin–cadmium (25°). All compositions established by analysis.

at room temperature and after a further annealing treatment of two weeks at 175°C, with agreeing results. The displacement liquid was ethylene dibromide, which has excellent wetting properties and only a small damping effect on the beam of the balance used for the experiments. The temperature dependence of the density of ethylene dibromide was determined both by the pyknometer method and the displacement method, using a standard aluminium specimen of known density. The temperature of the immersion liquid was controlled to $\frac{1}{10}$°C, and the temperature was measured by a copper–constantan thermocouple. All weighings were corrected for buoyancy, and the derived density values ρ_0, which are contained in table 1, were reduced to a common temperature

Table 1. Densities of Tin–Indium Alloys

(1)	(2)	(3)	(4)	(5)
0 (pure tin)	7·2854	7·2860	7·2858	7·2863
1·03	7·2862	7·2860	7·2861	7·2858
1·55	7·2837	7·2842	7·2840	7·2856
1·81	7·2833	7·2835	7·2834	7·2850
1·90	7·2689	7·2682	7·2685	7·2921
2·07	7·2813	7·2842	7·2828	7·2874
2·33	7·2847	7·2845	7·2846	7·2858
3·10	7·2809	7·2826	7·2818	7·2846
4·11	7·2822	7·2807	7·2815	7·2837
5·65	7·2801	7·2825	7·2813	7·2832

(1) Composition of alloy (at. % In); (2) density after 4 months at room temp. (g cm^{-3}, 25°c); (3) density after further annealing (g cm^{-3}, 25°c); (4) mean density (g cm^{-3}, 25°c); (5) theoretical density (g cm^{-3}, 25°c).

of 25°C. Table 1 also includes the theoretical densities, calculated from the measured lattice spacings by means of the expression $\rho_T = mMN/a^2c$ where M is the mean atomic weight, N is the number of atoms in the unit cell, and m is the atomic unit of mass ($1{\cdot}66035 \times 10^{-24}$ g); a and c are the lattice spacings in cm. In the critical composition range the observed and calculated densities differ appreciably, showing the development of vacant lattice sites. Figure 4 shows the percentage of lattice sites occupied as calculated from ρ_0 and ρ_T, and the number of valency electrons per site (i.e. electrons per unit cell / 4) assuming a valency of 4 for tin. From 1·8 to 2·4 atomic % In, the number of valency electrons per site is less than that corresponding to complete filling of atomic sites.

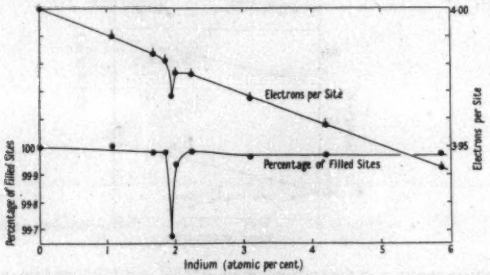


Figure 4. Tin–indium alloys : percentage of lattice sites occupied, and number of valency electrons per lattice site.

3.5. *The Systems Tin–Cadmium, Tin–Zinc and Tin–Mercury*

The maximum solubility of cadmium in tin is approximately 1·2 wt. % at 132·5°C. Seven alloys were annealed for three weeks at 155°C, and quenched. The lattice spacings are plotted in figure 3. As for tin–indium alloys, there is little change in c, but an overall decrease in a, with an increase in the axial ratio. The same type of anomaly as in the tin–indium alloys occurs, but at smaller solute concentrations.

Zinc and mercury dissolve in tin only to very limited extents. For completeness, however, two tin–zinc alloys and one tin–mercury alloy were investigated. The results are given in table 2. The results are insufficient to decide whether

Table 2

Alloy no.	Composition at. % Zn	Composition at. % Hg	a(kx) 25°C	c(kx) 25°C	c/a (25°C)
1	0·30	—	5·8194	3·1746	0·54552
2	0·70	—	5·8190	3·1746	0·54556
3	—	0·72	5·8169	3·1741	0·54561

anomalies similar to those in tin–indium and tin–cadmium alloys are present, but the fact that the a spacing for alloy 1 (0·3 atomic % Zn) lies on a straight line joining the a spacings of alloy 2 and of pure tin suggests that no anomaly of the magnitude of that in the tin–cadmium alloys at the same composition occurs.

§ 4. Discussion

Of the solute metals studied, only antimony and bismuth decrease the axial ratio of tin. Indium and cadmium ca se anomalies in the lattice spacing–composition curves which are consistent with the development of vacant sites. Figure 5 shows the axial ratios plotted against electron–atom ratio, assuming that the metals concerned exhibit their group valencies (Sb and Bi = 5, Sn = 4, In = 3, Cd, Zn and Hg = 2). Antimony and bismuth decrease the axial ratio to comparable extents as the electron–atom ratio increases, while elements of group valency less than 4 increase the axial ratio, and in the cases of indium and cadmium, anomalous variations occur as the electron–atom ratio is lowered.

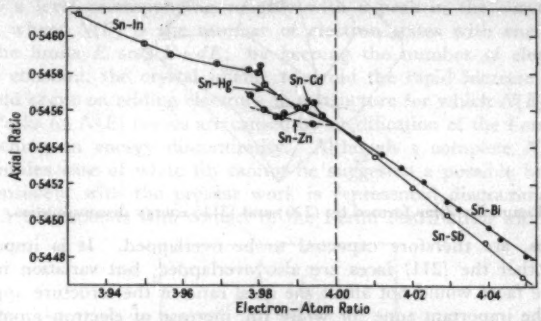


Figure 5. Variation of axial ratio with electron–atom ratio in tin-rich alloys.

Lead also decreases the axial ratio, but to a much smaller extent per atomic per cent than in the case of antimony and bismuth. These regularities suggest that the lattice spacings of tin-rich alloys are affected by overlaps of electrons from the first Brillouin zone, as previously discussed for magnesium (Raynor 1940, 1942, Hume-Rothery and Raynor 1940). If the overlap is in directions parallel to the *a* axes of the crystal, an increase in electron–atom ratio is expected to expand the *a* spacings relatively to the *c* spacings, giving a decrease in axial ratio, while a decrease in electron–atom ratio leads to the reverse effect.

It is of interest, therefore, to examine the Brillouin zone structure of white tin. The possible planes of energy discontinuity, together with the corresponding structure factor $|S|$, are:

	{200}	{101}	{220}	{211}	{112}
$\|S\|$	4	$2\sqrt{2}$	4	$2\sqrt{2}$	4

Mott and Jones (1936) suggest that the first Brillouin zone is bounded by the group of planes {101}. The volume of this zone corresponds with 4·11 electrons per atom; the inscribed sphere, however, corresponds with only 1·30 electrons per atom. Unless, therefore, the energy gap across the {101} planes is very large, overlap of electrons into the second zone occurs. Increase of the overlap would tend to displace the {101} zone boundaries towards the origin of reciprocal space, but would have very little effect on the axial ratio; on this basis the systematic variations of axial ratio with electron concentration cannot be interpreted. It is more probable that the form of the first Brillouin zone allows overlap across some plane parallel to the *c* axis.

The appropriate first Brillouin zone should thus contain approximately four electrons per atom, with the possibility of overlap across planes parallel to the *c* axis. The zone enclosed by the {220} and {211} planes (figure 6) contains approximately 4·4 electrons per atom. The inscribed sphere touches the {220} planes at 3·23 electrons per atom; these faces, assuming reasonable

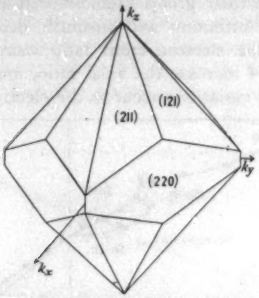


Figure 6. Zone formed by {220} and {211} energy discontinuities.

energy gaps, are therefore expected to be overlapped. It is impossible to decide whether the {211} faces are also overlapped, but variation in overlap across these faces would not affect the axial ratio of the structure appreciably. If this is the important zone for white tin, increase of electron–atom ratio by alloying with antimony and bismuth increases the {220} overlap, and causes a relative increase in the *a* spacings, with a reduction in axial ratio. Decrease of electron–atom ratio by alloying with indium, cadmium, zinc and mercury conversely contracts the *a* spacings, leading to an increase in axial ratio. Schubert (1950) has independently suggested a similar Brillouin zone for white tin in his discussion of the allotropy of tin.

Other planes of energy discontinuity lie within the volume of figure 6. These, derived from {200} and {101} reflections, are indicated in figure 7. Before reaching the {220} planes, therefore, the Fermi sphere passes through these discontinuities. The lattice spacing measurements already discussed

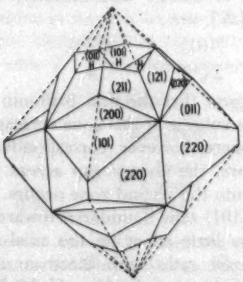


Figure 7. Planes of energy discontinuity within the zone of figure 6.

suggest that the {220} face is overlapped; a tentative explanation of the anomalous lattice spacings in tin–indium and tin–cadmium alloys is possible assuming that the {211} faces are closely approached by the Fermi distribution. As the electron–atom ratio increases on passing from the solubility limit of indium in tin towards pure tin, the number of electrons per lattice site also increases smoothly to a value of approximately 3·98 (figure 4). Further addition of tin causes lattice sites to be left vacant; the number of electrons per lattice site tends to remain constant† until the vacant sites disappear on the addition of a further small amount of tin. Omission of atoms to avoid increase in the number of electrons per lattice site is typical of cases where the Brillouin zone is filled to a level corresponding roughly with a peak in the curve of $N(E)$ against E, where $N(E)$ is the number of electron states with energies lying between the limits E and $E+dE$; by keeping the number of electrons per lattice site constant, the crystal is able to avoid the rapid increase in energy which would occur on adding electrons to a structure for which $N(E)$ is falling rapidly. Peaks on $N(E)$ curves are caused by modification of the Fermi surface on approaching an energy discontinuity. Although a complete $N(E)$ curve for the complex case of white tin cannot be suggested a possible high-energy portion consistent with the present work is represented diagrammatically in figure 8; A corresponds with contact of the Fermi distribution with the {220}

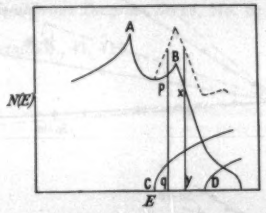


Figure 8. Possible $N(E)$ curve for white tin towards the upper limit of the Fermi distribution (diagrammatic).

planes, C corresponds with the onset of the overlap across these planes, and B corresponds with contact of the Fermi surface with the {211} planes. We suggest that for tin, the zone is filled to the level xy, and that overlap across {211} planes (point D) has not occurred; at the solubility limit of indium in tin, the zone is filled to some level such as pq. If B corresponds to 3·98 electrons per lattice site, the behaviour of the alloys of tin with solutes of lower valency may be interpreted, since further increase in the level to which the zone is filled corresponds with a falling $N(E)$ curve. This argument is supported by the relative slopes of the a spacing–composition curves for indium in tin on either side of the anomaly. As tin is added to the saturated solution of indium in tin, the a spacing increases, but the rate of increase falls off as the critical composition is approached; this is consistent with figure 8, since, while the $N(E)$ curve pB is rising steeply, a given increase in the number of electrons per site causes a progressively smaller increase in the number of overlapping electrons. Once the level of B is exceeded a given increase in the

† The sharp fall in the number of electrons per lattice site for the alloy containing 1·9 atomic % of indium is not understood.

number of electrons per site will cause a progressively greater increase in the number of overlapping electrons, and a more marked rate of expansion of a, as observed experimentally.

Although the zone structure of white tin may be discussed only qualitatively the lattice spacing effects in tin-rich alloys suggest that the appropriate Brillouin zone is bounded principally by the {220} and {211} families, possibly truncated by the re-entrant {101} faces H of figure 7. Once the Fermi distribution has become distorted by passing through planes of energy discontinuity within the main zone, it is difficult to predict at which outer planes overlap may occur; the experimental evidence however, suggests overlap from the {220} faces, and contact of the Fermi distribution with the {211} planes. The linearity of the a and c spacing curves for tin–antimony and tin–bismuth alloys suggests that no overlap across {211} planes occurs up to an electron–atom ratio of 4·08.

It is of interest to compare the mean atomic volumes of tin-rich alloys (i.e. volume of unit cell / 4) with the atomic volumes calculated additively from those of the pure components. Figure 9 shows that, for the alloys with antimony and bismuth, the atomic volumes fall well below those corresponding

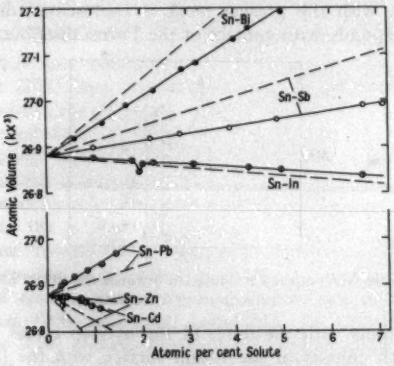


Figure 9. Atomic volumes of tin-rich alloys.

to an additive relationship between the atomic volumes of the components. In other cases the deviation is positive; the smallest deviation occurs for indium. In general, the deviation from additivity is the more marked the greater the difference between the atomic volumes of the solute and solvent metals, and, except for lead, solutes with larger atomic volumes than white tin cause negative deviations, while solutes of smaller atomic volume correspond with a positive deviation. Although lead has an atomic volume similar to that of antimony, the deviations in the tin–lead and tin–antimony alloys are in opposite senses.

In view of the difficulty of assessing atomic size influences for tin-rich alloys, no further discussion is justified at present; it is probable however that the major factor involved in the lattice distortion of tin by the metals studied is overlap of electrons from the appropriate Brillouin zone in directions at right angles to the tetragonal axis.

ACKNOWLEDGMENTS

The work described in the present paper forms part of a general programme in progress in the authors' laboratory, and grateful acknowledgment for generous financial assistance and other facilities which have aided the research is made to the Royal Society, the Chemical Society, Imperial Chemical Industries Ltd., the Tin Research Institute and the Department of Scientific and Industrial Research.

REFERENCES

AXON, H. J., and HUME-ROTHERY, W., 1948, *Proc. Roy. Soc.* A, **193**, 1.
BERRY, R. L. P., HENRY, W. G., and RAYNOR, G. V., 1951, *J. Inst. Metals*, **78**, 643.
EGERTON, A. C., and LEE, W. B., 1923, *Proc. Roy. Soc.* A, **103**, 487.
ERFLUIG, H. P., 1939, *Ann. Phys., Lpz.*, **34**, 136.
HUME-ROTHERY, W., LEWIN, G. F., and REYNOLDS, P. W., 1936, *Proc. Roy. Soc.* A, **157**, 167.
HUME-ROTHERY, W., and RAYNOR, G. V., 1940, *Proc. Roy. Soc.* A, **177**, 27.
IEVANS, A., STRAUMANIS, M., and KARLSONS, K., 1938, *Z. Phys. Chem.* B, **40**, 347.
MOTT, N. F., and JONES, H., 1936, *The Theory of the Properties of Metals and Alloys* (Oxford : Clarendon Press).
OWEN, E. A., 1943, *J. Inst. Metals*, **69**, 1.
OWEN, E. A., and ROBERTS, E. W., 1939, *Phil. Mag.* [7], **27**, 294.
RAYNOR, G. V., 1940, *Proc. Roy. Soc.* A, **174**, 457; 1942, *Ibid.*, **180**, 107; 1947, *Inst. Metals Annotated Equilibrium Diagram Series*, No. 6; 1949, *Trans. Faraday Soc.*, **45**, 698.
SCHUBERT, K., 1950, *Z. Metallkde.*, **41**, 423.

Taylor Instability of the Interface between Superposed Fluids —Solution by Successive Approximations

By R. L. INGRAHAM
Institute for Advanced Study, Princeton, New Jersey, U.S.A.

MS. received 25*th September* 1953

Abstract. A solution of the hydrodynamical equations for the case of Taylor instability for two fluids of constant density is sought by the method of successive approximations. The expansion parameter is the initial amplitude of the interface; the validity of the method then assumes convergence of the series involved. The linear equations to be solved at the pth step for the pth order unknown fields are derived. The solution is actually carried out up to the second approximation. An interesting feature of this solution is that its time growth comports not only the integral modes $\exp(\sigma t)$ and $\exp(2\sigma t)$ but also the irrational mode $\exp(\sqrt{2}\sigma t)$, in terms of the growth parameter

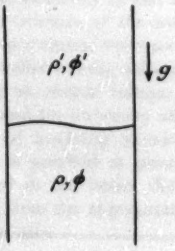
$$\sigma \equiv \{gk(\rho' - \rho)/(\rho' + \rho)\}^{1/2}.$$

A NATURAL way to attack the solution of non-linear field equations, such as those of hydrodynamics, is the method of successive approximations. The method is valid, and one can therefore get arbitrarily near to the exact solution, in a domain of the parameters for which the series involved converge. This method is here applied to the hydrodynamical equations with the boundary conditions corresponding to Taylor instability of two fluids of constant density.

First the general formulae are derived, that is, the linear equations to be solved at the pth step for the pth order unknown fields. A standard procedure for the solution of these sets of equations is suggested. Then the solution up to $p=2$ (good up to third order terms) is obtained.

This method was used by Rosenhead (1931) to obtain the solutions for Helmholtz instability good up to third order terms.

It has been argued on physical grounds that, although these series converge for $t=0$, they diverge for any positive time (and any value of the initial amplitude parameter) no matter how small. A mathematical proof of this, while perhaps possible on the basis of the explicit solutions here obtained, would seem to be difficult. In any case, it is suggested that the solution, good up to the first few approximations beyond the linear, serves to give a qualitative idea of the growth of the motion in the initial stages.



§1. The General Equations

The equations characterizing Taylor instability are, first

$$\nabla^2\phi=0, \qquad \nabla^2\phi'=0, \qquad \ldots\ldots(1.1)$$

primes characterizing the upper fluid, where our sign convention is that the velocities are the negative gradients of the respective velocity potentials:

$$\mathbf{u}=-\nabla\phi, \qquad \mathbf{u}'=-\nabla\phi'.$$

Second, where $y=\eta(x, t)$ is the interface, the kinematical boundary condition

$$\frac{DF}{Dt}\equiv\partial_t\eta-\partial_x\phi\partial_x\eta+\partial_y\phi=0;\ F\equiv\eta-y \qquad \ldots\ldots(1.2)$$

and

$$\frac{DF}{Dt}\equiv\partial_t\eta-\partial_x\phi'\partial_x\eta+\partial_y\phi'=0 \qquad \ldots\ldots(1.2)'$$

where ∂_t means $\partial/\partial t$, ∂_x means $\partial/\partial x$, ∂_y means $\partial/\partial y$. Equations (1.2), (1.2)′ are to hold at $y=\eta$. Third, pressure continuity across the interface

$$(\rho'-\rho)g\eta-\rho'\partial_t\phi'+\rho\partial_t\phi+\tfrac{1}{2}\rho'[(\partial_x\phi')^2+(\partial_y\phi')^2]-\tfrac{1}{2}\rho[(\partial_x\phi)^2+(\partial_y\phi)^2]=0 \quad \ldots(1.3)$$

which likewise is to hold at $y=\eta$. The arbitrary time functions are absorbed in ϕ, ϕ' in (1.3).

We shall solve these equations by the method of successive approximations, assuming the amplitude of the initial disturbance small, and expanding the various quantities into series in terms of it. For a given value of this amplitude parameter, the solutions will then be good for times such that these series converge. We shall first derive the general formulae, then use them to obtain the solution up to third order terms.

Let the expansions be

$$\phi=\sum_{p=1}^{\infty}\phi_p, \qquad \phi'=\sum_{p=1}^{\infty}\phi_p', \qquad \eta=\sum_{p=1}^{\infty}\eta_p \qquad \ldots\ldots(1.4)$$

where the subscript denotes the order of smallness. We are given $\eta_1=a_1\cos kx$ at $t=0$, and a_1 is the expansion parameter.

For any function f of y, we can express its value at $y=\eta$ in terms of its value and those of its derivatives at $y=0$:

$$f|_{y=\eta}=\sum_{\substack{i=0\\ r_1, r_2, \ldots r_i=1,\ldots\infty}}^{\infty}\frac{1}{i!}(\partial_{yy\ldots y}{}^{i}f)_{y=0}\eta_{r_1}\eta_{r_2}\cdots\eta_{r_i}.$$

Using this expansion, we can express equations (1.2), (1.2)′ and (1.3) in terms of quantities evaluated at $y=0$. When the series (1.4) are substituted into these equations, and the results expressed in terms of quantities evaluated at $y=0$, we get the set of equations which will be given shortly. First, it is convenient to define

$$U_p\equiv-\sum_{q=1}^{p}\frac{1}{i!}\partial_{yy\ldots y}{}^{i}(\partial_x\phi_q)|_{y=0}\sum_{\substack{i=0\\ r_1+\ldots+r_i=p-q}}\eta_{r_1}\eta_{r_2}\cdots\eta_{r_i} \qquad \ldots\ldots(1.5)$$

$$V_p\equiv-\sum_{q=1}^{p}\frac{1}{i!}\partial_{y\ldots y}{}^{i}(\partial_y\phi_q)|_{y=0}\sum_{\substack{i=0\\ r_1+\ldots+r_i=p-q}}\eta_{r_1}\eta_{r_2}\cdots\eta_{r_i} \qquad \ldots\ldots(1.6)$$

$$P_p\equiv\sum_{q=1}^{p}\frac{1}{i!}\partial_{y\ldots y}{}^{i}(\partial_t\phi_q)|_{y=0}\sum_{\substack{i=0\\ r_1+\ldots+r_i=p-q}}\eta_{r_1}\eta_{r_2}\cdots\eta_{r_i} \qquad \ldots\ldots(1.7)$$

and
$$U_p^{(-)} \equiv U_p + \partial_x \phi_p \qquad \ldots\ldots(1.5')$$
$$V_p^{(-)} \equiv V_p + \partial_y \phi_p \qquad \ldots\ldots(1.6)'$$
$$P_p^{(-)} \equiv P_p - \partial_t \phi_p \qquad \ldots\ldots(1.7)'$$

and P_p', ... $U_p'^{(-)}$ etc. the corresponding expressions in ϕ'. U_p is actually the pth order part of the x-velocity of the lower fluid evaluated at t, x and $y=\eta$, and a corresponding remark holds for V_p and P_p. $U_p^{(-)}$ is U_p less the term in the pth order part of ϕ, $-\partial_x \phi_p$. The notation means that for given p, for every q less than p we form the sum of the products of η_r's such that the subscripts in each summand add up to $p-q$, etc.

Then (1.2) and (1.2)′ break up into the equations

$$\partial_t \eta_p + \partial_y \phi_p = V_p^{(-)} - \sum_{r+q=p} U_q \partial_x \eta_r \qquad (\text{at } y=0) \qquad \ldots\ldots(1.8)$$
$$\partial_t \eta_p + \partial_y \phi_p' = V_p'^{(-)} - \sum_{r+q=p} U_q' \partial_x \eta_r \qquad (\text{at } y=0) \qquad \ldots\ldots(1.8)'$$

and (1.3) becomes

$$(\rho'-\rho) g \eta_p - \rho' \partial_t \phi_p' + \rho \partial_t \phi_p = \rho' P_p'^{(-)} - \rho P_p^{(-)} - \tfrac{1}{2}\rho' \sum_{q+q'=p} \{U_q' U_{q'} + V_q' V_{q'}\}$$
$$+ \tfrac{1}{2}\rho \sum_{q+q'=p} \{U_q U_{q'} + V_q V_{q'}\} \quad (\text{at } y=0). \qquad \ldots\ldots(1.9)$$

Finally, the linear equations (1.1) become simply

$$\nabla^2 \phi_p = 0, \qquad \nabla^2 \phi_p' = 0. \qquad \ldots\ldots(1.10)$$

At the pth step therefore one has to satisfy the sets (1.8), (1.8)′, (1.9), and (1.10) of *linear* equations in the unknowns ϕ_p, ϕ_p' and η_q. The right members are functions of the known ϕ_q, ϕ_q', and η_q for $q=1, \ldots p-1$. The standard method for doing this is as follows.

The x-dependence of η_p can usually be determined from (1.8), (1.8)′ by inspection. Then setting $\eta_p = \Phi_p(t)\Psi(x)$, where $\Phi_p(t)$ is the unknown time dependence, this is substituted into (1.8), (1.8)′. The x-dependence of ϕ_p, ϕ_p' is given by (1.8), (1.8)′ and therefore the general form of the y-dependence is determined by their harmonicity. (1.8), (1.8)′ give the values of the y-derivatives at $y=0$ and therefore normalize the constant factors. Then substituting these values of ϕ_p, ϕ_p' into (1.9), we get an ordinary second order differential equation for $\Phi_p(t)$. The solution is made unique by the demands

$$\Phi_p = \frac{d\Phi_p}{dt} = 0 \quad \text{at } t=0; \quad (p \geqslant 2). \qquad \ldots\ldots(1.11)$$

§ 2. Actual Solutions for $p=1$ and 2

$p=1$. This is the linear approximation, in which no non-linear terms make themselves felt. Equation (1.10) gives

$$\nabla^2 \phi_1 = \nabla^2 \phi_1' = 0. \qquad \ldots\ldots(2.1)$$

Equations (1.8), (1.8)′ give

$$\partial_t \eta_1 + \partial_y \phi_1 = 0 \qquad (\text{at } y=0). \qquad \ldots\ldots(2.2)$$
$$\partial_t \eta_1 + \partial_y \phi_1' = 0 \qquad (\text{at } y=0). \qquad \ldots\ldots(2.2)'$$

Equation (1.9) gives

$$(\rho'-\rho) g \eta_1 - \rho' \partial_t \phi_1' + \rho \partial_t \phi_1 = 0 \qquad (\text{at } y=0). \qquad \ldots\ldots(2.3)$$

Write $\eta_1 = \Phi_1(t)\cos kx$ and substitute into (2.2), (2.2)′. We get, in virtue of (2.1)

$$\left.\begin{aligned} \phi_1 &= -\frac{e^{ky}}{k}\partial_t\Phi_1\cos kx + f_1(t) \\ \phi_1' &= -\frac{e^{-ky}}{-k}\partial_t\Phi_1\cos kx + f_1'(t). \end{aligned}\right\} \quad \ldots\ldots(2.4)$$

Put these into (2.3)

$$(\rho'-\rho)g\Phi_1\cos kx - \frac{\rho'}{k}\partial_t^2\Phi_1\cos kx - \rho'\partial_t f_1' - \frac{\rho}{k}\partial_t^2\Phi_1\cos kx + \rho\partial_t f_1 = 0. \quad \ldots(2.5)$$

We can take $f_1 = f_1' = 0$. Then we get

$$\partial_t^2\Phi_1 - \sigma^2\Phi = 0, \qquad \sigma^2 \equiv gk\frac{\rho'-\rho}{\rho'+\rho}. \quad \ldots\ldots(2.6)$$

The solution of this desired is $\Phi = a_1\cosh\sigma t$, because we impose the boundary conditions

$$\Phi_1 = 1, \qquad \partial_t\Phi_1 = 0 \qquad \text{at } t = 0.$$

σ is real under the assumption $\rho' > \rho$, leading to instability. Substituting back, we get the first order solutions

$$\left.\begin{aligned} \phi_1 &= -\frac{a_1\sigma}{k}e^{ky}\cos kx\sinh\sigma t \\ \phi_1' &= \frac{a_1\sigma}{k}e^{-ky}\cos kx\sinh\sigma t \\ \eta_1 &= a_1\cosh\sigma t\cos kx. \end{aligned}\right\} \quad \ldots\ldots(2.7)$$

$\underline{p=2}$. When now the solutions (2.7) are put into the equations (1.8), (1.8)′ and (1.9) for $p=2$, and the following identities used to convert to functions of the double angle:

$$\cos^2 kx - \sin^2 kx = \cos 2kx$$
$$\cos^2 kx = \tfrac{1}{2}(1+\cos 2kx)$$
$$\cosh^2\sigma t = \tfrac{1}{2}(\cosh 2\sigma t + 1), \qquad \sinh^2\sigma t = \tfrac{1}{2}(\cosh 2\sigma t - 1),$$

the linear equations to be solved in the second approximation take the following form:

$$\partial_t\eta_2 + \partial_y\phi_2 = \tfrac{1}{2}a_1^2k\sigma\sinh 2\sigma t\cos 2kx \qquad (\text{at } y=0) \quad \ldots\ldots(2.8)$$

$$\partial_t\eta_2 + \partial_y\phi_2' = -\tfrac{1}{2}a_1^2k\sigma\sinh 2\sigma t\cos 2kx \qquad (\text{at } y=0) \quad \ldots\ldots(2.8)'$$

$$(\rho'-\rho)g\eta_2 - \rho'\partial_t\phi_2' + \rho\partial_t\phi_2 = \tfrac{1}{4}(\rho-\rho')a_1^2\sigma^2(1+\cosh 2\sigma t)\cos 2kx + \tfrac{1}{2}(\rho-\rho')a_1^2\sigma^2\cosh 2\sigma t \qquad (\text{at } y=0) \quad \ldots\ldots(2.9)$$

$$\nabla^2\phi_2 = \nabla^2\phi_2' = 0. \quad \ldots\ldots(2.10)$$

We write $\eta_2 = \Phi_2(t)\cos 2kx$ and substitute this into (2.8), (2.8)′. The y-dependence of ϕ_2, ϕ_2' must be e^{2ky}, e^{-2ky} respectively, and these equations then fix the factors, giving

$$\phi_2 = \frac{e^{2ky}}{2k}(-\partial_t\Phi_2 + \tfrac{1}{2}a_1^2k\sigma\sinh 2\sigma t)\cos 2kx + f_2(t)$$

$$\phi_2' = \frac{e^{-2ky}}{2k}(\partial_t\Phi_2 + \tfrac{1}{2}a_1^2k\sigma\sinh 2\sigma t)\cos 2kx + f_2'(t).$$

When these are put into (2.9), we find first that, as to the pure time-dependent parts, we can take $f_2=f_2'=\frac{1}{4}a_1^2\sigma\sinh 2\sigma t$ and second, that the remaining differential equation for Φ_2 becomes

$$\partial_{tt}^2\Phi_2-2\sigma^2\Phi_2=\tfrac{1}{2}g^{-1}\sigma^4a_1^2(1-\cosh 2\sigma t).$$

The general solution of this is

$$\Phi_2=C\cosh\sqrt{2}\sigma t+D\sinh\sqrt{2}\sigma t-\tfrac{1}{2}g^{-1}a_1^2\sigma^2(1+\cosh 2\sigma t).$$

The boundary conditions (1.11) then fix the constants C, D as

$$D=0, \qquad C=a_1^2\sigma^2/2g.$$

Hence, the second order solutions are

$$\left.\begin{aligned}
\phi_2&=\frac{1}{4}\frac{a_1^2\sigma}{gk}\{e^{2ky}(-\sqrt{2}\sigma^2\sinh\sqrt{2}\sigma t\\
&\qquad+(\sigma^2+kg)\sinh 2\sigma t)\cos 2kx+kg\sinh 2\sigma t\}\\
\phi_2'&=\frac{1}{4}\frac{a_1^2\sigma}{gk}\{e^{-2ky}(\sqrt{2}\sigma^2\sinh\sqrt{2}\sigma t\\
&\qquad+(gk-\sigma^2)\sinh 2\sigma t)\cos 2kx+kg\sinh 2\sigma t\}\\
\eta_2&=\frac{a_1^2\sigma^2}{2g}\{\cosh\sqrt{2}\sigma t-\tfrac{1}{2}(1+\cosh 2\sigma t)\}\cos 2kx.
\end{aligned}\right\}\qquad(2.11)$$

It is interesting to note that the solutions correct up to third order terms, given by (2.7) and (2.11), involve not only the rate-of-growth parameters σ and 2σ, integral multiples of σ, as might have been expected, but also the irrational multiple $\sqrt{2}\sigma$.

Reference

Rosenhead, L., 1931, *Proc. Roy. Soc.* A, **134**, 170.

The Approach to the Steady State in Gaseous Thermal Diffusion and its Application to Determining the Dependence of Gas Diffusivity on the Concentration Ratio

By P. T. NETTLEY†
Physics Department, University of Birmingham

Communicated by T. L. Ibbs; MS. received 21st May 1954

Abstract. The approach to the steady state in gaseous thermal diffusion has been studied using a conventional apparatus consisting of two vessels joined by a connecting tube. The investigation was confined to mixtures of hydrogen and nitrogen. The rate of approach to the steady state was found to be exponential with mixtures of these two gases. It was also found that the rate of approach to the steady state decreased slightly as the percentage of the lighter gas increased. Since the theory shows that the rate of approach to equilibrium is proportional to the coefficient of diffusion D_{12}, this suggests that the coefficient D_{12} decreases as the percentage of light gas increases. The experimental variation of the coefficient D_{12} with concentration as determined in this way is compared with that predicted for the rigid elastic sphere type of molecule. The advantages of this method of measuring changes of the coefficient of diffusion with concentration ratio over the ordinary diffusion method are discussed.

§ 1. Introduction

The application of a temperature gradient to a uniform mixture of two gases produces a separation of the constituent gases due to thermal diffusion. The extent to which this separation proceeds is limited by the opposing effect of ordinary diffusion. When the two effects balance a steady state is obtained. The manner in which this equilibrium state is approached has been investigated by Chapman (1919), and later by Clark Jones and Furry (1946). A slightly modified account of this latter treatment is given by Grew and Ibbs (1952).

Clark Jones and Furry discuss the manner in which equilibrium is approached in a simple apparatus consisting of two vessels maintained at different temperatures and joined by a connecting tube. In this treatment it is assumed that the temperature gradient along the connecting tube is uniform, that the volume of the vessels is large compared with that of the connecting tube, and that the coefficient of diffusion of the gas mixture in the apparatus varies as the square of the temperature. The theory then shows that for two vessels at temperatures T and T' $(T'>T)$, the variation of the separation S with time t is given by the equation

$$S=S_{\text{equil}}(1-e^{-bt}), \qquad \ldots\ldots(1)$$

where S_{equil} is the separation when the steady state is attained. The approach to the steady state is therefore exponential. The relaxation time, that is the

† Now with the United Kingdom Atomic Energy Authority, Capenhurst.

time required for S to attain $1-e^{-1}$ of its final value, is

$$\tau = \frac{1}{b} = \frac{NN'}{N+N'} \frac{L}{nD_{12}} \frac{T}{\sigma \Delta T} \ln\left(\frac{T'}{T}\right), \qquad \ldots\ldots(2)$$

where N and N' are the number of molecules in the vessels at temperature T and T' respectively, n is the number density of both species of molecules in the gas mixture at temperature T, L and σ are the length and cross-sectional area of the connecting tube respectively and $\Delta T = T' - T$. The relaxation time is therefore proportional to the length of the connecting tube, and inversely proportional to its cross-sectional area. τ depends on the pressure only through the coefficient D_{12}; as this is inversely proportional to the pressure, τ varies directly with pressure. Since τ is dependent on the coefficient D_{12}, measurements of the rate of approach to equilibrium for different mixtures of two gases should provide information on the variation of D_{12} with the concentration of the mixture.

Dr. T. L. Ibbs had shown in connection with early experiments on thermal diffusion that the approach to equilibrium was approximately exponential. As this work had not been published, Dr. Ibbs suggested that a more complete investigation could now profitably be made.

§ 2. Experimental Method

2.1. *Apparatus*

The apparatus and procedure are essentially those used by Ibbs (1925) in his early experiments on thermal diffusion. A diagram of the apparatus is given in figure 1. A cylindrical brass vessel A surrounded by a steam jacket formed

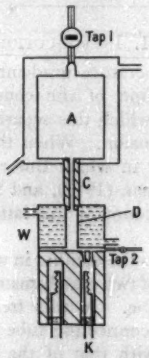


Figure 1. The diffusion apparatus.

the hot side of the apparatus. This vessel was joined to the cold side of the apparatus by a connecting tube C. The cold side of the apparatus consisted of a much larger tube D, the open cell K of a katharometer and the intermediate volume above the katharometer.

The volumes of the hot and cold sides were 82·5 cm^3 and 7·6 cm^3 respectively. The connecting tube was 5·0 cm long and 0·315 cm diameter. The dimensions of the connecting tube were chosen to make the rate of approach to the steady

state reasonably long whereas in the previous work by Ibbs a wide and short connecting tube was used to obtain equilibrium as quickly as possible. In the work described here the time required to reach the steady state was about thirty minutes.

The Shakespear katharometer (see Daynes 1920, 1933) was used to analyse the gas mixture under investigation and also to follow the thermal separation on the cold side of the apparatus. A constant stream of water through the water jacket W maintained the cold side at constant temperature and also served to protect the katharometer from the influence of the steam bath.

2.2. *Experimental Procedure*

Since the rate of approach to the steady state is dependent on both the temperature gradient and the pressure, it was essential that the temperatures of the hot and cold sides should be steady both before the thermal separation commenced and throughout the separation process. Therefore the flow method developed by Ibbs was used. This method fills the apparatus with a uniform mixture at closely the temperatures of the different parts before the separation process begins.

The gases under investigation were mixed under pressure in cylinders and stored to attain uniform composition. The mixture was then passed into the apparatus at tap 1 and out at tap 2. Tap 2 was connected to a simple water bubbler which indicated the magnitude of the flow. When the apparatus was full of the mixture under investigation, as indicated by the steadiness of the katharometer reading, the composition of the mixture was determined. Still maintaining a steady gas flow, steam was passed through the steam jacket. The flow of the mixture opposed the tendency for thermal separation; a steady flow was maintained so that this separating effect was overcome and gas of uniform composition filled the whole apparatus. The gas stream attained the temperature of the containing vessel as it passed through and the rate of flow of gas up to a reasonable speed did not affect the katharometer zero.

When conditions were steady, as demonstrated by the katharometer, the gas flow was stopped by closing taps 1 and 2. Separation then commenced due to thermal diffusion rapidly at first, and then more slowly until equilibrium was reached. The variation of the separation with time was determined by measuring the concentration ratio of the gas mixture in the cold side every half minute by means of the katharometer.

§ 3. Results

In the experiments the rate of approach to the steady state on the cold side was measured by means of the katharometer. The assumption is made that conditions in the apparatus were those which Clark Jones and Furry term quasi-stationary. In this case the rate of approach to the steady state in the hot side of the apparatus is identical with that in the cold side. Clark Jones and Furry show that such conditions apply if the volume of each vessel is much larger than the connecting tube as is the case in this work.

The variation of the cold side separation S_c with the time thermal diffusion had progressed is shown in figure 2 for a gas mixture containing 60·3% hydrogen

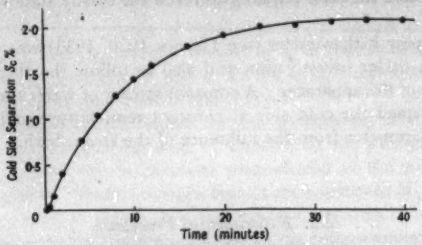


Figure 2. Variation of cold side separation S_c with time for a mixture containing 60·3% hydrogen and 39·7% nitrogen.

and 39·7% nitrogen. Similar curves were obtained for other mixtures of hydrogen and nitrogen.

The variation of $\log(S_{c(equll)} - S_c)$ with time for three mixtures of hydrogen and nitrogen is shown in figure 3. The results for all other mixtures investigated

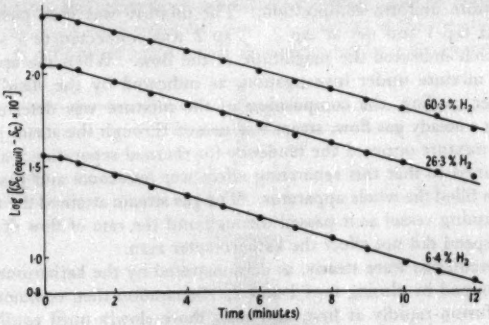


Figure 3. Variation of log $(S_{c(equll)} - S_c)$ with time.

were similar. Figure 3 shows that the separation follows an exponential variation with time except for a small time, of the order of half a minute, at the start of the thermal diffusion process. This initial time lag was due to the fact that communication between the open cell of the katharometer and the remainder of the cold side was established by three small holes in the katharometer block. There was therefore a small time lag before any change in composition in the main part of the cold side was transmitted to the katharometer due to the time taken for gas to diffuse through the holes. Daynes (1933) has shown that in cases such as this, where the katharometer time lag was much shorter than the relaxation time for thermal diffusion, the variation of composition with time in the katharometer cell is the same as the variation in the volume above the

cell; but the actual composition in the cell lags a few seconds behind that in the volume above the cell. Therefore if the separation obeys the relation

$$S = S_{\text{equil}}(1 - e^{-bt}),$$

then the separation measured by the katharometer is

$$S = S_{\text{equil}}[1 - \exp\{-b(t - t_0)\}],$$

where t_0 is the katharometer time lag.

The linear relation obtained between $\ln(S_{c(\text{equil})} - S_c)$ and t proves the physical validity of the theory of Clark Jones and Furry. The slope of the straight line is b and is the reciprocal of the relaxation time τ. For a given apparatus with a constant temperature gradient, τ is dependent only on the coefficient D_{12} as shown in equation (2). The values of τ obtained for different mixtures of hydrogen and nitrogen are given in table 1. These values are corrected to standard conditions; namely $T' = 373°\text{K}$, $T = 288°\text{K}$ and pressure = 760 mm Hg.

Table 1. Variation of Relaxation Time τ with Concentration

% Hydrogen	6·4	13·7	18·3	26·3	30·5	39·8	54·5	60·3	73·2	82·7
τ (minutes)	7·30	7·53	7·52	7·66	7·63	7·76	7·72	8·23	8·18	8·13

Equation (2) can be expanded to give

$$\tau = \frac{T^2}{D_{12}} V' V \left(\frac{V}{T} + \frac{V'}{T'}\right)^{-1} \frac{L}{\sigma \Delta T} \ln\left(\frac{T'}{T}\right) \qquad \ldots\ldots(3)$$

where V' and V are the volumes of the hot and cold vessels respectively.

Von Obermayer (1880) obtained a value of 0·740 cm² sec^{-1} for the coefficient of diffusion of hydrogen and nitrogen at 288°K and 760 mm Hg pressure. This result was obtained by letting the pure components diffuse into each other and is therefore a mean diffusion coefficient over the whole concentration range. This value can be compared with the present results by substituting the mean value of τ, 7·7 minutes, into equation (3) with the other appropriate values. The value obtained for D_{12} is 0·77 cm² sec^{-1} at 288°K and 760 mm Hg pressure. This compares quite well with von Obermayer's value.

§ 4. Discussion of Results

4.1. *The Coefficient of Diffusion* D_{12}

The theory of Clark Jones and Furry contains certain simplifying assumptions. The main assumption is that the coefficient D_{12} is proportional to T^2, whereas von Obermayer (1880) has shown that in the case of hydrogen and nitrogen it is proportional to $T^{1.74}$. Nevertheless there is surprisingly good agreement between the value of D_{12} obtained by the thermal diffusion method and the accepted value obtained from ordinary diffusion methods. This, together with the fact that the approach to equilibrium has been shown to be exponential as predicted by theory, shows that the analysis of Clark Jones and Furry gives the correct general physical picture of the approach to equilibrium in gaseous thermal diffusion.

4.2. *The Variation of D_{12} with Concentration Ratio*

Although the thermal diffusion method may be unsuitable for making absolute determinations of the coefficient D_{12} due to simplifying approximation in the theory, the method is, however, a powerful one for determining relative values of D_{12} of gas mixtures having the same temperature variation of D_{12}. In this case errors caused by approximations in the theory affect the values of the coefficient D_{12} of each mixture to the same extent. Relative values of the coefficient D_{12} of the mixtures of hydrogen and nitrogen that were investigated have been calculated from their respective relaxation times and are given in table 2.

Table 2. Variation of D_{12} for Hydrogen and Nitrogen with Concentration

% Hydrogen	D_{12} (observed relative values) $cm^2\ sec^{-1}$	$[D_{12}]_2$ (calculated r.e.s.)
0	—	0·784
6·4	0·81 ± 0·02	0·784
13·7	0·79 ± 0·02	0·783
18·3	0·79 ± 0·01	0·783
26·3	0·78 ± 0·01	0·781
30·5	0·78 ± 0·01	0·779
39·8	0·77 ± 0·01	0·776
54·4	0·77 ± 0·02	0·770
60·3	0·73 ± 0·02	0·767
73·2	0·73 ± 0·02	0·758
82·7	0·74 ± 0·03	0·751
100	—	0·729

The results show that the coefficient D_{12} decreases slightly as the percentage of lighter gas is increased. The errors shown result from the probable error in the measurement of the relaxation time. This error varies from mixture to mixture since both the magnitude of the thermal separation and the sensitivity of the katharometer depend upon the composition of the mixture.

Chapman (1917) has obtained formulae for the first approximation to the coefficient of diffusion $[D_{12}]_1$ for several molecular models. Whatever model is used the first approximation is independent of the concentration of the mixture in which diffusion occurs. Chapman has shown that the second approximation $[D_{12}]_2$ is of the form

$$[D_{12}]_2 = [D_{12}]_1/(1-\Delta),$$

where Δ is a function depending on the concentration of the mixture and the intermolecular forces.

For comparison, values of $[D_{12}]_2$ have been calculated using Chapman's formula for Δ. The calculation of Δ has been made on the assumption that the molecules are rigid elastic spheres and that their radii are those found from viscosity measurements. The first approximation $[D_{12}]_1$ has been chosen so that the value of $[D_{12}]_2$ is equal to the observed value of D_{12} for the mixture approaching nearest to an equimolecular mixture. These calculated values of $[D_{12}]_2$ are given in table 2.

Further approximations to the coefficient D_{12} show a larger variation with composition than does $[D_{12}]_2$; but the variation of D_{12} with composition for molecules acting as point centres of repulsive force obeying an inverse power

law ($F(r)=kr^{-\nu}$) is less than for rigid elastic spheres. It is known from viscosity data (cf. Chapman and Cowling 1939) that gases fit the inverse power law quite well over limited ranges of temperature if ν, the force index, is chosen correctly. For hydrogen $\nu=11{\cdot}3$ and for nitrogen $\nu=8{\cdot}8$ in the temperature range 288–373°K. The variation of $D_{12}/[D_{12}]_1$ with ν has been evaluated by Chapman in the special case of a Lorentzian gas in which the ratio m_1/m_2 of the molecular masses in the mixture is very large and the ratio n_1/n_2 of the molecular number densities is very small. This variation of $D_{12}/[D_{12}]_1$ is given in table 3. In the other extreme case where $m_1/m_2\to\infty$ and $n_1/n_2\to\infty$, $D_{12}/[D_{12}]_1\to 1$. The Lorentz case is an extreme case which gives the largest variation of D_{12} with concentration ratio n_1/n_2; as the ratio m_1/m_2 decreases so does the variation of $D_{12}/[D_{12}]_1$ with n_1/n_2.

Table 3. Variation of $D_{12}/[D_{12}]_1$ and $[D_{12}]_2/[D_{12}]_1$ with ν in the Lorentz Case

ν	∞ (r.c.s.)	17	13	9	5	3	2
$[D_{12}]_2/[D_{12}]_1$	1·083	1·049	1·039	1·023	1·000	1·125	3·250
$D_{12}/[D_{12}]_1$	1·132	1·072	1·056	1·031	1·000	1·132	3·396

Table 3 shows that, in the Lorentz case, molecules such as hydrogen and nitrogen which have values of ν between 9 and 13 should have a smaller variation of D_{12} with concentration ratio than the variation of $[D_{12}]_2$ calculated on the assumption that the molecules are rigid elastic spheres. The experimental results in table 2 show however that for hydrogen–nitrogen mixtures the variation of D_{12} with concentration ratio is somewhat larger than the variation of $[D_{12}]_2$ calculated on the assumption that the molecules are rigid elastic spheres. The results however are in general agreement with theory as regards the order of magnitude and the character of the variation.

The strongest previous evidence that has been obtained for the variation of the coefficient of diffusion with composition is provided by work performed at Halle (see Lonius 1909). In this work hydrogen–oxygen, hydrogen–carbon dioxide, helium–argon and hydrogen–nitrogen mixtures were investigated. In the case of the first three pairs of gases it was found that the coefficient D_{12} decreased with the proportion of the lighter gas as was later predicted by Chapman's theory. The magnitude of the variations observed was also slightly greater than the variation of $[D_{12}]_2$ calculated on the assumption that the molecules are rigid elastic spheres. It was found at Halle that the value of the coefficient D_{12} for hydrogen–nitrogen mixtures increased when the proportion of hydrogen increased; but only two mixtures were investigated. This is the reverse of that which has been found in the present work and also of that which is predicted by Chapman's theory.

The work at Halle was performed by observing the mutual diffusion of the two gases in the mixture. Such observations are difficult to make and are liable to fairly large experimental error. Also, in such experiments the composition of the gas mixture in different parts of the diffusion column varies considerably during the diffusion process. This makes the interpretation of the results difficult for showing how the coefficient D_{12} varies with concentration ratio.

The method of using the rate of approach to the steady state in thermal diffusion does not suffer from this difficulty. The maximum change of concentration during the thermal diffusion process is quite small. During the

experiments described here the maximum variation in concentration was just over two per cent on the cold side and much less on the hot side. This enabled the coefficient D_{12} to be measured over a small range of concentrations.

§ 5. Conclusions

The use of the rate of approach to the steady state in thermal diffusion as a method of determining the variation of the coefficient of diffusion with concentration ratio appears superior to previous ordinary diffusion methods. The lack of data on this subject suggested that further work could profitably be undertaken with other gas mixtures. These investigations are being made by Mr. P. M. Rolph.

Acknowledgments

The author wishes to thank Dr. T. L. Ibbs for suggesting the problem and also for his valuable help and advice. He is also indebted to the Department of Scientific and Industrial Research for the award of a research grant.

References

Chapman, S., 1917, *Phil. Trans. Roy. Soc.* A, **217,** 115 ; 1919, *Phil. Mag.*, **38,** 185.

Chapman, S., and Cowling, T. G., 1939, *The Mathematical Theory of Non-Uniform Gases* (Cambridge : University Press).

Clark Jones, R. C., and Furry, W. H., 1946, *Rev. Mod. Phys.*, **18,** 163.

Daynes, H. A., 1920, *Proc. Roy. Soc.* A, **97,** 273; 1933, *Gas Analysis by Measurement of Thermal Conductivity* (Cambridge : University Press).

Grew, K. E., and Ibbs, T. L., 1952, *Thermal Diffusion in Gases* (Cambridge : University Press).

Ibbs, T. L., 1925, *Proc. Roy. Soc.* A, **107,** 470.

Lonius, A., 1909, *Ann. Phys., Lpz.*, **29,** 664.

von Obermayer, J. V. A. E., 1880, *Wien. Ber.*, **81,** 1102.

Optical and Photo-Electrical Properties of Indium Antimonide

By D. G. AVERY, D. W. GOODWIN, W. D. LAWSON and T. S. MOSS†
Radar Research Establishment, Great Malvern, Worcs.

Communicated by G. G. Macfarlane; MS. received 24th May 1954

Abstract. Measurements have been made of the refractive index and absorption index of InSb by reflection methods with polarized radiation. Extensive studies of absorption constant have been made by direct transmission measurements.

Both photovoltaic and photoconductive effects have been observed in InSb. At room temperature the photosensitivity extends to approximately 7·5 μ, a considerably longer wavelength than for any other known material.

§ 1. Introduction

In view of the similarity of the compounds of groups III and V of the periodic table to the well-known semiconducting elements of group IV (Welker 1952), it was thought that InSb, by comparison with gray tin, might have a very low activation energy. It has also been suggested previously that arsenides and antimonides were particularly hopeful materials to examine with a view to increasing the present wavelength limits of photoconductors (Moss 1950). Some transmission measurements on InSb have already been reported in a recent note by Tanenbaum and Briggs (1953). The experimental data described here are discussed by Moss (1954).

§ 2. Reflection Measurements

The method of measurement was that of Avery (1952), the wavelength range covered being from the visible region to 3·2 μ, the limit of the PbS cell detector used. The measurements were made on a cleavage face of a single crystal. The sample was prepared by fusing high purity elements in stoichiometric proportions in an evacuated silica tube. The crystal was then grown from the melt by slow cooling to room temperature. Measurements of conductivity and Hall constant showed that this crystal was relatively imp re, having $1{\cdot}5 \times 10^{18}$ n-type carriers at room temperature. The radiation was provided by a double monochromator with NaCl prisms, the resolution being 0·1 μ.

The results obtained for the real and imaginary parts of the dielectric constant i.e. $n^2(1-k^2)$ and $2n^2k$ where the complex refractive index is $n(1-ik)$, are shown in figure 1. For wavelengths greater than 1 μ the real part is constant at 14, corresponding to a refractive index of 3·75.

† Now at Royal Aircraft Establishment, Farnborough.

The absorption index is approximately constant from 1μ to 3μ, and falls rapidly at slightly longer wavelengths, indicating an absorption edge. The estimated errors in the absorption determination are indicated on the graph.

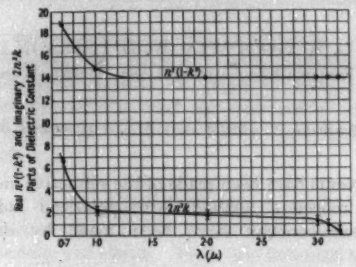


Figure 1. Optical constants of impure InSb.

§ 3. Other Optical Measurements

3.1. *Transmission Measurements*

For the impure specimen referred to above, where the absorption edge lay at relatively short wavelengths, LiF prisms were used in the double monochromator. For all other specimens NaCl prisms were used. Radiation was chopped at 5 c/s and a Schwarz thermopile with KRS 5 window was used as detector. Readings were taken over the required wavelength range with the specimen in the beam, and then taken over the same range with the specimen removed. When an absorption cell with windows was used, these windows were left in the beam for the measurement with the specimen removed.

In calculating the absorption constant allowance was made for the reflectivity calculated from the refractive index, and for any loss caused because the whole of the radiation did not pass through the aperture in the backing plate holding the sample. The samples were usually large enough to cover a hole 3 mm by 2 mm. Thicknesses ranged from 0·09 mm to 0·8 mm.

The absorption cell was used when measurements above or below room temperature were undertaken. It was made in the form of a demountable Dewar flask fitted with periclase (MgO) windows. The specimen, mounted on a brass backing plate, was bolted to a solid brass plate which was kept in direct contact with the heating or cooling liquid by use of a copper–glass seal joined to the inner part of the Dewar flask.

3.2. *Impure Material*

The complete room temperature absorption curve for the impure sample is shown by the full line of figure 2. For wavelengths less than $3{\cdot}1\mu$ reflection data on the absorption index have been used to find K, the absorption coefficient. The steep part of the absorption edge was measured with LiF prisms in the monochromator. No significant difference in the K values resulted on changing the resolution from $0{\cdot}06\mu$ to $0{\cdot}03\mu$, so that the finite slope of the edge is a property of the material and not determined by instrumental resolution. Taking the

absorption edge as being the point of maximum slope (i.e. at $K \sim 2000\,\text{cm}^{-1}$) we find the optical activation energy to be 0·38 ev.

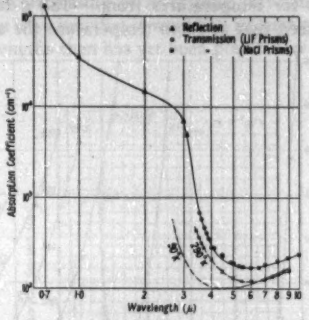


Figure 2. Absorption spectrum of impure InSb.

The two broken curves were taken with the absorption cell. NaCl prisms were used, the resolution being 0·2 μ. The difference between the 290°K curve and the full line is attributed partly to the different prisms and resolutions used, and partly to inhomogeneities in the specimen showing up under the different focusing arrangements used in the two measurements. The two broken curves were taken under identical conditions and thus give the temperature shift of the edge. The direction of shift is of increasing energy gap on cooling (i.e. as for germanium) and the magnitude of the shift at an absorption level of 400 cm^{-1} is $dE/dT = -4{\cdot}3 \times 10^{-4}\,\text{ev deg}^{-1}$.

3.2. *Pure Material*

A sample of highly purified InSb was obtained from S.E.R.L., Baldock, and a wide range of transmission measurements made with it. The material was not a single crystal, but the crystallites were large, i.e. ~1 mm^3 in volume.

A check was made of the temperature dependence of the conductivity of this material. The results, plotted as log (conductivity) against reciprocal temperature in figure 3, show that the material is an intrinsic semi-conductor down to 220°K.

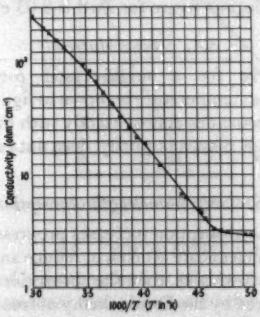


Figure 3. Conductivity of pure InSb.

The slope of the straight line is typical of that found by other workers (e.g. Welker 1953, Weiss 1953, Cunnell *et al.* 1953), namely 0·49 ev.

Absorption curves for temperatures from −183°c to 226°c are shown in figure 4. It will be seen that at room temperature the absorption edge lies at

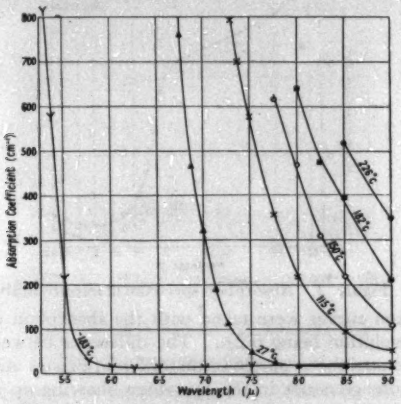


Figure 4. Spectral absorption of pure InSb.

much longer wavelengths than for the impure sample shown in figure 2. All the curves have finite slope, so that it is not possible to give unique values for the absorption edges. Taking an absorption level of 400 cm^{-1} for the 'edge', the following figures are found for its position:

Temperature (°c)	−183	+27	115	150	182	226
Wavelength (μ)	5·45	6·93	7·70	8·09	8·46	8·84

Although the absorption edge does not seem very steep when plotted against wavelength as in figure 4, it is actually very steep in terms of energies. At room temperature K changes by e:1 for an energy change of only 0·0069 ev (i.e. only $\sim\frac{1}{4}kT$). For comparison, to change the absorption constant for Ge by e:1 at the steepest part of the absorption curve requires an energy change of 0·017 ev (Moss 1952), and the analogous figure for ZnS is 0·03 ev (Piper 1953).

3.4. p-*type Material*

The absorption curve for a single-crystal p-type sample containing $2{\cdot}5 \times 10^{17}$ holes/cm^3 at room temperature is shown in figure 5. It will be observed that the absorption edge coincides almost exactly with that found for the intrinsic material, i.e. the addition of p-type impurity does not cause a shift of the edge as n-type impurity does.

3.5. *Long Wavelength Absorption*

For n-type material the absorption increases progressively with wavelength at long wavelengths. If sufficient care is taken to reduce any frequency-independent part of the absorption to a low value, then the absorption coefficient is found to be proportional to λ^2 as required by the Drude theory of free carrier absorption. The results plotted as K against λ^2 are shown in figure 6.

As may be seen from figure 5, the p-type material does not show the expected increase of absorption at long wavelength. There are indications that the intrinsic InSb shows absorption increasing as λ^2 at long wavelengths, although in this material the absorption is so low that it would be necessary to carry the measurements to longer wavelengths than has yet been possible in order to obtain precise data.

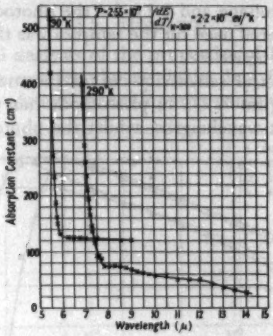


Figure 5. Absorption spectrum of p-type InSb.

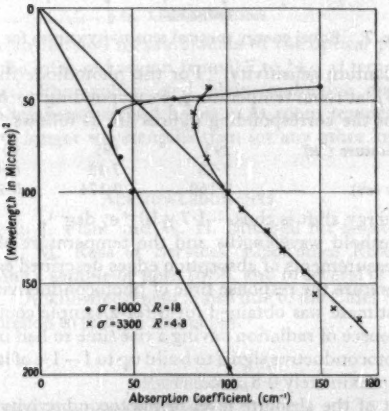


Figure 6. Long wavelength absorption of n-type InSb.

§ 4. Photoelectric Effects

Both photovoltaic and photoconductive effects have been observed in pure InSb. At room temperature the photosensitivity extends to considerably longer wavelengths than for any other known material.

Measurements of photovoltage were made at the contact between a tungsten probe and a cleavage face of a crystal. For the photoconductive experiments a thin rectangular bar (0·18 mm thick) with soldered end connections was used. All measurements were made with radiation chopped at 500–1000 c/s, with a matching transformer, high-gain amplifier, and homodyne detector to reduce the bandwidth to $\frac{1}{2}$ or $\frac{1}{16}$ c/s as required to give adequate signal–noise ratios. A double monochromator with NaCl prisms provided the radiation, the optical bandwidth being 0·2 μ for the photovoltage and 0·1 μ for the photoconductivity, at a wavelength of 7·5 μ. The material used for the samples was the high purity type with characteristics as shown in figure 3.

The equal-energy spectral sensitivity curves (normalized to 100% at their maxima) are shown in figure 7. The photodiode has approximately constant energy sensitivity at short wavelengths, whilst the photoconductor shows more

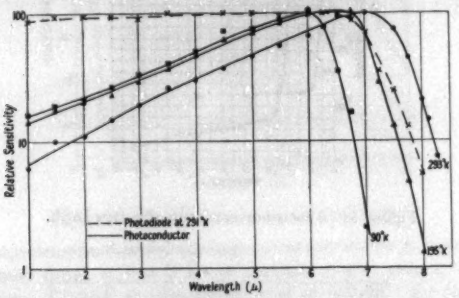


Figure 7. Equal energy spectral sensitivity curves for InSb.

nearly constant quantum sensitivity. For the photodiode the threshold wavelength is $\lambda_{1/2}=7{\cdot}15\,\mu$ at room temperature, corresponding to $h\nu=0{\cdot}174$ ev. For the photoconductor the corresponding figures are as follows:

Temperature (°K)	293	195	90
$\lambda_{1/2}$ (μ)	7·75	7·12	6·40
Energy (ev)	0·160	0·174	0·194

The average energy shift is thus $-1{\cdot}7\times10^{-4}$ ev deg^{-1}.

The above threshold wavelengths and the temperature shift are in general agreement with measurements of absorption edges described earlier.

At room temperature the response time of photoconductivity was too rapid to measure, but an estimate was obtained for it for a sample cooled with liquid air. With a spark-gap source of radiation having a rise time to half intensity of 0·3 μsec the time for the photoconductive signal to build up to $1-1/e$ of its maximum values was found to be approximately 0·8 μsec.

A measurement of the absolute level of photoconductivity was made. The source was a 24-watt tungsten filament lamp at 1 metre from the specimen. Hence the density of quanta incident on the specimen was $2{\cdot}2\times10^{15}$ cm^{-2} sec^{-1} allowance being made for 33% reflection at the InSb–air interface (as calculated from the optical constants of figure 1) and the fact that only 88% of the total quanta lie in the acceptable wavelength range, the effective number of quanta absorbed per second was $1{\cdot}3\times10^{15}$ cm^{-2}. The photoconductor was matched into the amplifier

by a transformer of 250:1 voltage step-up, and a steady voltage of 0·3 volt was applied in series with the photoconductor and transformer primary. Under these conditions the observed r.m.s. signal voltage at the transformer primary was $9{\cdot}6 \times 10^{-9}$ volt. For this specimen there was no evidence of any current noise, although, as in all the photoelectric measurements, the low signal voltages necessitated considerable care in reducing extraneous interference.

§ 5. Mobility

For use in theoretical analysis of the properties of InSb, mobility values are required. Figure 8 summarizes the available data for electrons, and shows considerable dependence of mobility on impurity concentration.

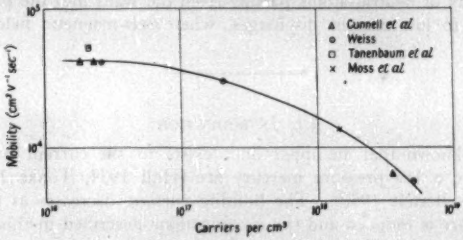


Figure 8. Dependence of electron mobility on carrier concentration.

§ 6. Conclusions

Reflection and transmission measurements of the optical properties of InSb have been made in the infra-red region from 0·7 to 14 μ at temperatures between −183°C and 226°C. Photovoltaic and photoconductive measurements have been made at room temperature and below. At room temperature the photosensitivity extends to longer wavelengths than for any other known material.

Acknowledgments

Thanks are due to J. Plant and W. H. Mitchell for assistance in preparing specimens, to Mr. I. M. Ross of Services' Electronics Research Laboratory, Baldock, for providing a sample of pure InSb, and to several of our colleagues for helpful discussions. Acknowledgment is also due to the Chief Scientist, Ministry of Supply, for permission to publish this paper.

References

Avery, D. G., 1952, *Proc. Phys. Soc.* B, **65**, 425.
Cunnell, F. A., Saker, E. W., and Edmond, J. T., 1953, *Proc. Phys. Soc.* B, **66**, 1115.
Moss, T. S., 1950, *Proc. Phys. Soc.* B, **63**, 167; 1952, *Photoconductivity in the Elements* (London: Butterworths); 1954, *Proc. Phys. Soc.* B, **67**, in the press.
Piper, W. W., 1953, *Phys. Rev.*, **92**, 23.
Tanenbaum, M., and Briggs, H. B., 1953, *Phys. Rev.*, **91**, 1561.
Weiss, H., 1953, *Z. Naturf.*, **8a**, 463.
Welker, H., 1952, *Z. Naturf.*, **7a**, 744; 1953, *Ibid.*, **8a**, 248.

Current Limitation in the Low-Pressure Mercury Arc

By J. E. ALLEN and P. C. THONEMANN
Atomic Energy Research Establishment, Harwell, Berks.

MS. received 12*th July* 1954

Abstract. Measurements on current limitation in the low-pressure mercury arc are reported and a theory is given which depends on the limit to the ion generation set by the flux of neutral atoms passing from the walls into the plasma. The paper refers to low current discharges, where self-magnetic field effects are negligible.

§ 1. Introduction

It is well known that an upper limit exists to the current which can be carried by a low-pressure mercury arc (Hull 1934, Tonks 1937, Alfvén 1950, von Bertele 1950). The limiting current decreases as the mercury vapour pressure is reduced and the investigations described in this paper were confined to a study of the phenomenon at low pressures ($\sim 10^{-4}$ mm Hg) and currents (~ 10 A), in order to avoid complications introduced by the self-magnetic field. In this pressure range the mean free path for collisions between positive ions and neutral atoms is greater than the tube radius, so that most of the ions produced fall freely to the tube wall without colliding with neutral mercury atoms.

It is shown that a simple interpretation of the current limitation can be given on the basis that the flux of positive ions to the tube wall cannot exceed the flux of neutral atoms leaving the wall (probe measurements have shown that these fluxes become comparable as the tube current is increased). As the limiting current is approached most of the mercury atoms are ionized during their first transit across the tube and the neutral gas density at the centre of the tube is considerably reduced. Thus the degree of ionization becomes much higher at the tube axis than near the wall.

§ 2. Apparatus

Preliminary measurements (previously unpublished) were made by Thonemann and Jeffries using a discharge tube similar to that described by Thonemann and Cowhig (1951). In these early experiments, however, the condensed mercury temperature was appreciably higher than the temperature of the cooling bath and the pumping action of a liquid air trap reduced the vapour pressure at the anode end of the tube. The later measurements were made using an anode bulb containing condensed mercury, so that the vapour pressure at the anode was accurately known (figure 1). Thermostatically controlled refrigerator units maintained the cathode and anode baths at the same temperature. The discharge tubes, which were of Pyrex glass about 1 metre long and 56 mm internal diameter,

were continuously evacuated by a two-stage mercury diffusion pump and the residual permanent gas pressure was about 2×10^{-7} mm Hg while the discharge was running.

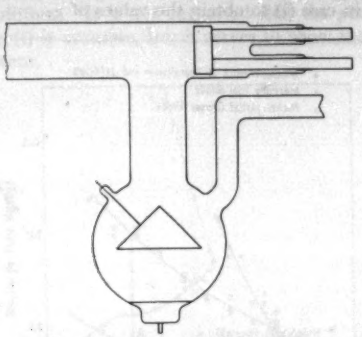


Figure 1. The water-cooled nickel-plated copper anode and adjacent mercury pool. The bulb and half the neck are immersed in a cooling bath.

§ 3. Experimental Results

At a given vapour pressure the limiting current was reached either by reducing the resistance in series with the tube or by raising the supply voltage. For convenience, the former method was generally used. The observed maximum currents are plotted in figure 2. In figure 3 they are plotted as a function of ν_0, the number of neutral atoms leaving unit area of tube wall per second at the end of the tube where the lowest vapour pressure existed. The values of ν_0 were calculated assuming that ν_0 was unaffected by the discharge and corresponded to the flux of neutral atoms leaving the mercury surface (except in the early experiments, when a liquid air trap was used).

(i) *Results of Thonemann and Jeffries*

The volt–ampere characteristics obtained by Thonemann and Jeffries were similar to curve A in figure 4. When the constant current region of the characteristic was reached, a relatively non-luminous region appeared at the anode and increased in length as the tube voltage was increased. Its appearance was a light reddish-brown in contrast to the blue-green of the normal plasma and spectroscopic examination showed the presence of mercury spark lines, some of which were identified as transitions in Hg^{3+}. In plotting the results in figure 3 the condensed mercury temperature was obtained by estimating the temperature drop across the glass wall of the cathode bulb, assuming a cathode drop of 10 volts. The vapour pressures were obtained from the data given by Ditchburn and Gilmour (1941), and the pumping effect of the liquid air trap was taken into account in calculating ν_0 at the anode, using the usual thermal effusion relations.

(ii) *Results obtained using the Mercury Pool as Anode*

Curve A in figure 4 is a typical volt–ampere characteristic. In this case the anomalous region developed at the cathode end of the tube, where the lowest

pressure existed. The condensed mercury at the anode end was at a higher temperature than that at the cathode end. Condensed mercury temperatures were estimated as in case (i) to obtain the values of ν_0.

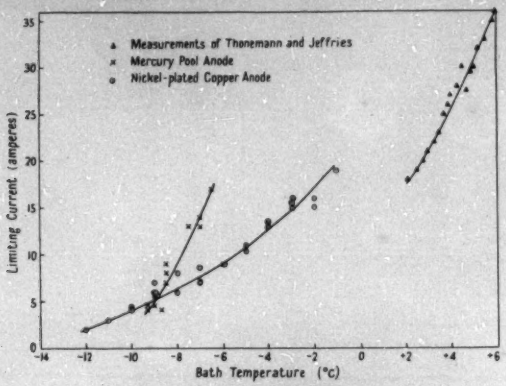


Figure 2. Observed limiting currents.

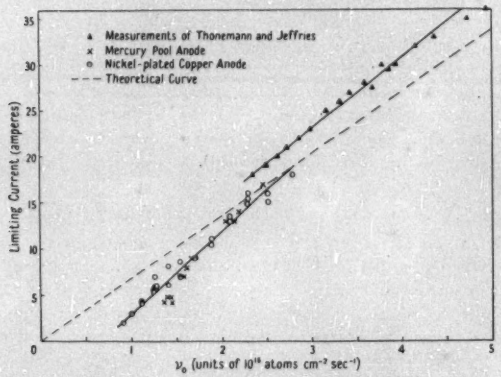


Figure 3. Limiting current as a function of ν_0.

(iii) *Results obtained using Nickel-Plated Copper Anode*

Curve B in figure 4 is a typical volt–ampere characteristic. The decrease in current after the maximum was reached was accompanied by a flickering of the discharge. Sometimes an anomalous region was observed at currents 'beyond' the maximum, i.e. at high tube voltages, and currents less than the maximum. The region has been observed at the anode and also in the middle of the tube. ν_0 at the anode was calculated from the condensed mercury temperature, which was equal to the temperature of the anode cooling bath.

(iv) *Comparison of Results*

Figure 3 shows reasonable agreement between the maximum currents obtained under different conditions. It is not suggested that the procedure adopted to determine v_0 in case (i) is accurate, but it serves to show that the different sets of results are consistent.

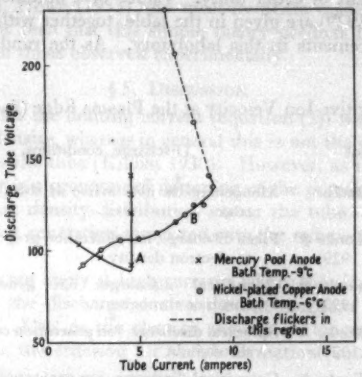


Figure 4. Voltage-current characteristics. The 'anomalous' region accompanies the constant current portion of curve A.

§ 4. Theory

(i) *The Plasma–Sheath Transition*

The low-pressure positive column can be conveniently divided into two regions (*a*) the plasma, in which the concentrations of positive and negative charges are high and closely equal, and (*b*) the positive ion space charge sheath which covers the tube wall.

If the positive ions arrive at the sheath edge with an energy e_2V_0 their density in the sheath at a plane at potential V is given by

$$n_2 = n_{2S}(1 - V/V_0)^{-1/2} \qquad \ldots\ldots(1)$$

where n_{2S} is the density at the sheath edge. If the distribution of electron velocity components normal to the sheath edge is Maxwellian and relatively few electrons reach the wall, the electron density in the neighbourhood of $V = 0$ is given by

$$n_1 = n_{1S} \exp(-e_1V/\boldsymbol{k}T_1) \qquad \ldots\ldots(2)$$

where e_1 is the electronic charge and T_1 is the electron temperature. At the plasma edge $n_{1S} = n_{2S} = n_S$, assuming that the ions are singly charged, and $dn_1/dx = dn_2/dx$ to a high degree of approximation so that $e_2V_0 = \frac{1}{2}\boldsymbol{k}T_1$ using equations (1) and (2). The positive ion velocity on arrival at the sheath is therefore given by

$$v_2 = (\boldsymbol{k}T_1/m_2)^{1/2} \qquad \ldots\ldots(3)$$

which is the criterion for sheath formation obtained by Bohm (1949). In a discharge tube ions are generated throughout the plasma and arrive at the plasma

boundary with a distribution of energies, furthermore the plasma merges continuously into the sheath, thus equation (3) is only approximate and will be replaced by

$$\bar{v}_2 = \alpha(kT_1/m_2)^{1/2} \qquad \ldots\ldots(4)$$

where α is a constant of order unity. Values of α obtained from the Tonks–Langmuir theory (1929) are given in the table, together with the value obtained from probe measurements in this laboratory. As the random current density

The Mean Positive Ion Velocity at the Plasma Edge ($\bar{v}_2 = \alpha(kT_1/m_2)^{1/2}$)

α	Reference	Discharge conditions	I_{2W}/I_{1R} (for Hg)
1·000	Theoretical (Bohm 1949)	Monoenergetic ions arrive at the plasma edge	$4{\cdot}128\times10^{-3}$
1·144	Theoretical (Tonks & Langmuir 1929)	Plane discharge, ion generation proportional to electron density	$4{\cdot}726\times10^{-3}$
1·179	Theoretical (Tonks & Langmuir 1929)	Cylindrical discharge, ion generation constant throughout	$4{\cdot}868\times10^{-3}$
1·199	Theoretical (Tonks & Langmuir 1929)	Spherical discharge, ion generation constant throughout	$4{\cdot}955\times10^{-3}$
1·213	Theoretical (Tonks & Langmuir 1929)	Cylindrical discharge, ion generation proportional to electron density	$5{\cdot}008\times10^{-3}$
1·227	Theoretical (Tonks & Langmuir 1929)	Plane discharge, ion generation constant throughout	$5{\cdot}067\times10^{-3}$
1·4	Experimental value from probe measurements (unpublished)	Low pressure mercury arc, ~5 amperes	$5{\cdot}8\times10^{-3}$

at the sheath edge is given by $I_{1R} = n_S|e_1|(kT_1/2\pi m_1)^{1/2}$ and the positive ion wall current is given by $I_{2W} = n_S e_2 \alpha (kT_1/m_2)^{1/2}$, the relation between these quantities is:

$$I_{1R} = (m_2/2\pi m_1)^{1/2} I_{2W}/\alpha. \qquad \ldots\ldots(5)$$

(ii) *Simplified Theory of Current Limitation in the Low-Pressure Mercury Arc*

Using equation (5), the tube current can be written

$$J_a = \left(\frac{\beta}{\alpha}\right)\left(\frac{m_2}{2\pi m_1}\right)^{1/2} I_{2W}\pi r_0^2 \qquad \ldots\ldots(6)$$

where β is the ratio of drift to random electron current, r_0 is the tube radius and radial variations of current density have been neglected. Equation (6) shows that if β remains approximately constant as the limiting current is approached, the tube current increases in proportion to I_{2W}, the positive ion wall current. Clearly, however, the flux of ions to the tube wall cannot exceed the flux of neutral atoms leaving so that as $I_{2W}/e_2 \to \nu_0$ the tube current approaches a limiting value given by

$$J_L = \left(\frac{\beta}{\alpha}\right)\left(\frac{m_2}{2\pi m_1}\right)^{1/2} \nu_0 e_2 \pi r_0^2 \qquad \ldots\ldots(7)$$

where ν_0 is the flux of neutral atoms leaving the wall. The dotted line† in figure 3 is a plot of equation (7) for $\alpha = 1{\cdot}4$ (the experimental value given in the table) and $\beta = 1$. The drift current is comparable with the random electron current (i.e. $\beta \sim 1$) in the low-pressure mercury arc, as shown by the probe measurements of Langmuir and Mott-Smith (1924), Killian (1930), Klarfeld (1938), Thonemann and Cowhig (1951), Howe (1953) and measurements made in this laboratory (unpublished). It is seen that this simple theory predicts limiting currents in good agreement with those observed experimentally.

§ 5. Discussion

The expression for the limiting current (equation (7)) was derived assuming a constant current density, whereas in general this is not the case, as the electron density varies across the tube (Killian 1930). However, as the limiting current is approached, there is a pronounced flattening of the potential distribution and hence of the electron density distribution across the tube. This results from the decrease in the ion generation associated with the reduced neutral gas density at the axis.

The theory does not apply at high currents where the electron density varies considerably across the discharge due to the self-magnetic field (Thonemann and Cowhig 1951). Also, as the electron density no longer follows the usual Boltzmann equation, the criterion for sheath formation (equation (4)) does not hold at high currents.

As the limiting current is approached, the neutral gas density at the tube axis is greatly reduced, due to ionization, and the degree of ionization at the centre of the tube approaches 100%. Near the wall, however, the degree of ionization remains quite small because $n_{2S}/n_{0S} \rightarrow (2T_0/\pi T_1)^{1/2}/\alpha$ as $I_{2W}/e_2 \rightarrow \nu_0$ (using equation (4)) which is only a few per cent for electron temperatures of the order of several volts.

A complete theory would include the calculation of β by formulating the energy and momentum balance equations for the electrons, but this is impossible at present as sufficient data are not available. For example the cross sections for ionization of the 6^3P_0, 6^3P_1 and 6^3P_2 states have not yet been determined and so the relative importance of multi-stage ionization cannot be assessed. A complete theory would also include the determination of the electron velocity distribution, which is far from spherically symmetrical, and calculations referring to rates of ionization and excitation must be based upon this anisotropic distribution. Collisions of the second kind take place and the velocity distribution should be used in the calculation of the de-excitation of the 6^3P_0, 6^3P_1 and 6^3P_2 states by electron impact. The population of these states depends also on the absorption and re-emission of radiation. Also, the relevant differential cross sections are required for a rigorous solution, as in all gas discharge problems (Massey and Burhop 1952). Finally, the loss of momentum and energy associated with electrons hitting the wall is important and would be included in a comprehensive theory.

Acknowledgment

The authors wish to thank Mr. C. J. F. Chaundy for assistance in the experimental work.

† When expressed in terms of the saturated vapour pressure the slope of this line corresponds approximately to 4 A cm^{-2} micron^{-1}.

References

Alfvén, H., 1950, *Cosmical Electrodynamics* (Oxford : Clarendon Press), p. 71.

von Bertele, H., 1950, *Proc. Instn Elect. Engrs*, Pt. II, **97**, 663.

Bohm, D., 1949, *The Characteristics of Electrical Discharges in Magnetic Fields* (New York: McGraw-Hill), Ch. 3.

Ditchburn, R. W., and Gilmour, J. C., 1941, *Rev. Mod. Phys.*, **13**, 310.

Howe, R. M., 1953, *J. Appl. Phys.*, **24**, 881.

Hull, A. W., 1934, *Elect. Engng*, **53**, 1435.

Killian, T. J., 1930, *Phys. Rev.*, **35**, 1238.

Klarfeld, B., 1938, *J. Tech. Phys., Moscow*, **5**, 913.

Langmuir, I., and Mott-Smith, H., 1924, *Gen. Elect. Rev.*, **27**, 538.

Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford : Clarendon Press), p. 14.

Thonemann, P. C., and Cowhig, W. T., 1951, *Proc. Phys. Soc.* B, **64**, 345.

Tonks, L., 1937, *Trans. Electrochem. Soc.*, **72**, 167.

Tonks, L., and Langmuir, I., 1929, *Phys. Rev.*, **33**, 1070; **34**, 876.

The Interpretation of the Properties of Indium Antimonide

By T. S. MOSS
Radio Department, Royal Aircraft Establishment, Farnborough, Hants.

MS. received 2nd June 1954

Abstract. The data given on the optical properties of InSb are analysed and precise values for the position and temperature dependence of the absorption edge are given.

The variation of the position of the absorption edge with impurity concentration is explained by the very low effective mass of the conduction electrons, which is estimated by three methods to be about 0·03 of the free electron mass.

§ 1. Introduction

As has been shown by the measurements reported by Avery, Goodwin, Lawson and Moss (1954), and also by the work of Tanenbaum and Briggs (1953), InSb shows a marked dependence of the position of the optical absorption edge on the degree of n-type impurity present in the specimens. This material is also remarkable for the extremely high mobility of its free electrons.

Both these phenomena are explained by a model in which the energy surfaces in momentum space of the bottom of the conduction band show much greater curvature than is encountered in other materials. Such a model has been proposed independently by Burstein (1954).

It follows from this model that (i) the effective mass of the electrons is small, and hence the mobility high, (ii) the density of states in the conduction band is small, so that a relatively low density of conduction electrons can fill the states to an appreciable depth and thus affect the absorption edge, and (iii) that the shift of the absorption edge with temperature will in part be determined by this process of filling up the lower levels on the conduction band.

From the inter-relation of these three phenomena, two estimates have been made of the effective electron mass. A third estimate has been made from the magnitude of the free carrier absorption at long wavelengths.

§ 2. The Absorption Edge and its Temperature Dependence

The curves of absorption coefficient K against wavelength for pure InSb given by Avery *et al.* do not all fall at the same rate, as is illustrated more clearly in figure 1, where $\log K$ is plotted against energy. It will be seen that at high absorption levels the points lie on straight lines, i.e. the absorption coefficient decreases exponentially with increasing quantum energy. The shift of the absorption edge with temperature is thus clearly dependent on the absorption level at which the edge is defined. On extrapolating the lines of figure 1 it is found that those for room temperature and above all intersect in the same region, at approximately 16 000 cm^{-1}, so that if the plots were in fact straight lines up to this level then at $K = 16\,000\ cm^{-1}$ there would be no shift with temperature at all.

Figure 2 shows the position of the absorption edge (for various K values) as a function of temperature, from which it may be seen that the quantum energy varies linearly with temperature, so that the shift dE/dT is constant with temperature at any given absorptive level, where E is the activation energy.

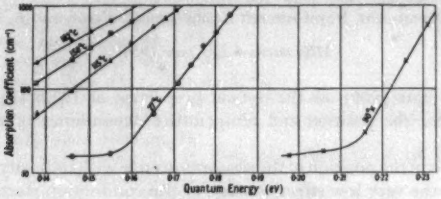


Figure 1.

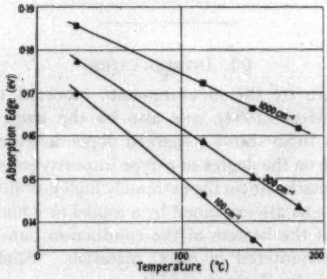


Figure 2. Temperature shift of absorption edge.

From the analysis of the data it is found that (above room temperature) the shift in 10^{-4} ev/°c is

$$dE/dT = -5{\cdot}3 + 1{\cdot}25 \log_{10} K = 1{\cdot}25 \log_{10} (K/16\,000) \qquad \ldots\ldots(1)$$

and the position of the absorption edge (in ev) at T°c is given by

$$E = 0{\cdot}205 + 1{\cdot}25(T+100)10^{-4} \log_{10} (K/16\,000). \qquad \ldots\ldots(2)$$

Below room temperature (i.e. at -183°c) the slope of the log (absorption) curve does not increase much. This is presumably because the apparent steepness is no longer due to the properties of the InSb, but due to instrumental limitations, i.e. the resolution of the spectrometer is now the limitation. It may be noted that at the short wavelength limit of transmission at -183°c the resolution in terms of energy was only half as good as at the corresponding point at room temperature.

From the complete absorption curve given in figure 2 of the accompanying paper (Avery *et al.*) it was found that maximum steepness of the absorption curve occurred at about 2000 cm^{-1}, and this point is identified with the true absorption edge. We may assume that at this level the effect of cooling will be to shift this point to shorter wavelengths and simultaneously make the curve pivot about this point. In other words we will assume that at the 2000 cm^{-1} level the observed shift is a true shift of absorption edge, there being no contribution due to the

blurring or broadening of the edge. It is found by extrapolating in figure 1 or by substituting $K = 2000$ in equations (1) and (2) that at 300°K the edge lies at 0·19 ev, that the shift above room temperature is $-1{\cdot}1 \times 10^{-4}$ ev/°C and the shift below room temperature is $-2{\cdot}2 \times 10^{-4}$ ev/°C.

§ 3. Theory

3.1. *Calculation of Effective Mass*

X-ray measurements† of the lattice constant for the impure n-type material and the pure material shown by Avery *et al.* yield the value 6·478 Å for both materials. Thus the marked difference in the positions of the absorption edges must be a fundamental property of the InSb, and cannot be attributed to a difference in crystal structure or to expansion of the lattice by impurities.

Measurements by Tanenbaum and Maita (1953) of Hall constant and conductivity have shown that to reconcile the values of instrinsic carrier concentrations calculated from these measurements with the activation energy appropriate to the room temperature absorption edge at 0·18 ev (the figure used by Tanenbaum—it is very near the estimate of 0·19 ev made in the present work) it is necessary to use effective masses m_e and m_h of electrons and holes such that $m_e m_h = 0{\cdot}007 m^2$. Use of $m_e m_h \sim m^2$ as is commonly assumed for semiconductors gives an activation energy of 0·37 ev, and furthermore if used with Tanenbaum and Maita's data would lead to an activation energy which *increased* with temperature, whereas $m_e m_h = 0{\cdot}007 m^2$ gives the shift in the direction observed in the absorption work. Between 300°K and 500°K the calculated shift from Tanenbaum and Maita's data is $dE/dT = -310^{-4}$ ev/°C. Thus either m_e or m_h must be extremely small.

Another feature of InSb is the remarkably high mobility of the electrons, about 42 000 $cm^2 v^{-1} sec^{-1}$ at room temperature. For holes the value is not unusual being approximately 600 $cm^2 v^{-1} sec^{-1}$. As the mobility increases with decreasing effective mass, it is probable that m_e has a low value while m_h/m is perhaps not much less than unity.

Magnetoresistance measurements by Pearson and Tanenbaum (1953) indicate that the energy zones in InSb are spherical, so that there should be no variation of effective mass with crystal directions. Hence for an electron at the bottom of the conduction band with effective mass m_e the relation between the energy E and wave vector $\mathbf{k}$ is

$$E = h^2\mathbf{k}^2/2m_e, \qquad \ldots\ldots(3)$$

hence

$$dE/d\mathbf{k} = h^2\mathbf{k}/m_e. \qquad \ldots\ldots(4)$$

Now the density of allowed states in $\mathbf{k}$ space is simply $dn/d\mathbf{k} = 2(4\pi\mathbf{k}^2)$ thus

$$dn/dE = 8\pi m_e \mathbf{k}/h^2 = (8\pi/h^3)(2m_e{}^3E)^{1/2}.$$

Integrating,

$$N = \int_0^{\Delta E} dn = \frac{8\pi m_e{}^{3/2}}{h^3}\,\frac{2\sqrt{2}}{3}\,(\Delta E)^{3/2}.$$

This expression gives N, the total number of states in an energy interval ΔE above the bottom of the conductor band. Putting in the appropriate constants we have

$$m_e/m = 4 \times 10^{-15} N^{2/3}/\Delta E. \qquad \ldots\ldots(5)$$

Now from the experimental data the room temperature absorption edge lies at 0·19 ev (at $K = 2000$ cm^{-1}) in pure material, and at 0·38 ev in n-type material

† Carried out by Miss A. Rennie of R.R.E. Malvern.

with $1{\cdot}5\times10^{18}$ carriers/cm^3. Postulating that the presence of these electrons is responsible for filling up the lower levels of the conduction band to the extent of $\Delta E=0{\cdot}38-0{\cdot}19=0{\cdot}19$ ev, and that the above equation is applicable we have†

$$m_e/m=4\times10^{-15}(1{\cdot}5\times10^{18})^{2/3}/0{\cdot}19=0{\cdot}028.$$

This simple theory thus explains the shift of the absorption edge with carrier concentration, in terms of a very small electron mass. Such a prediction is in general agreement with the low value of m_em_h found by Tanenbaum and Maita (1953) and the exceptionally high electron mobility in InSb.

As an outcome of this theory it would be expected that although n-type samples have the absorption edge at much shorter wavelengths than intrinsic samples, p-type samples by nature of the much larger effective mass of holes and consequent small curvature of the energy surface in **k** space should differ little from intrinsic samples in the position of the absorption edge. That this is so is shown by the absorption data for a p-type sample (with $2{\cdot}5\times10^{17}$ holes at room temperature) shown by Avery *et al.* The curve coincides almost exactly with that for an intrinsic sample at the same temperature, thus confirming the theory put forward to explain the shift of absorption edge with concentration in the n-type specimens, and the marked difference in the effective masses of holes and electrons.

3.2. *Calculation of the Effect of Filling of the Conduction Band on dE/dT*

This theory may now be used to estimate the temperature shift of the energy gap due to thermally filling the lower levels of the conduction band by intrinsic electrons. The effect of emptying the upper levels of the full band will be ignored as the higher effective mass of the holes means less curvature of the energy surface in **k** space and consequently a denser distribution of allowed levels.

Note that this effect does not necessitate *complete* occupancy of the lower conduction levels, but only that the density of states corresponding to the absorption constant considered (i.e. $K=2000$ cm^{-1}) should occur at a higher energy. Nevertheless, the state of occupancy is quite high above room temperature even in intrinsic InSb as the energy gap is less than $7\mathbf{k}T$ and the high value of m_h/m_e raises the Fermi level $1{\cdot}5\mathbf{k}T$ above the middle of the forbidden zone to within $2\mathbf{k}T$ of the bottom of the conduction band.

Rewriting equation (5) in its relative form

$$N_1^{2/3}/E_1=N_2^{2/3}/E_2=\text{constant}, \qquad \ldots\ldots(5\,a)$$

and using the previous values $\Delta E_1=0{\cdot}38-0{\cdot}19=0{\cdot}19$ ev when $N_1=1{\cdot}5\times10^{18}$, we have

$$N^{2/3}/\Delta E=7\times10^{12}. \qquad \ldots\ldots(6)$$

Hence at room temperature (27°c) for an intrinsic sample where $N=2{\cdot}2\times10^{16}$, $\Delta E_{27}=0{\cdot}01$ ev. At $T=182$°c, where interpolating from Tanenbaum and Maita's results gives $N_T=2{\cdot}1\times10^{17}$, we have $\Delta E_T=0{\cdot}054$ ev. We therefore assume that the filling of the lower levels of the conduction band has the effect of *increasing* the energy gap by the amount of ΔE given above. In the absence of this effect therefore the energy gap at room temperature (at $K=2000$ cm^{-1}) would be $E_{27}=0{\cdot}19-0{\cdot}01=0{\cdot}18$ ev and at 182°c, $E_{182}=0{\cdot}172-0{\cdot}054=0{\cdot}12$ ev.

From these two figures the average shift above room temperature would be $dE/dT=-4{\cdot}0\times10^{-4}$ ev/°c. It should be noted that value is in good agreement

† It should be noted that in this example the Fermi level will be well *above* the bottom of the conduction band and the electron gas will be highly degenerate.

with the shift observed in the impure specimen quoted by Avery *et al.* (namely $-4{\cdot}3\times10^{-4}$ ev/°c) where the effect of intrinsic electrons should be negligible. We thus conclude that the true temperature shift of the energy gap is near $dE/dT=-4{\cdot}0\times10^{-4}$ ev/°c. This value also approximates to that deduced by Tanenbaum and Maita from their analysis of Hall constant data.

3.3. *Calculation of Effective Mass from Temperature Dependence of the Energy Gap*

The shift of the absorption edge with temperature is caused mainly by two effects, namely (see Moss 1952):

(i) A broadening of the allowed bands of levels due to interaction with lattice vibrations. As the bands always broaden on heating, $(dE/dT)_1$ for this effect is always negative. This effect has been treated by Fan (1951) who gives the following expression:

$$(dE/dT)_1=4\pi\left\{\frac{128\omega^2k^3}{243\pi}\right\}^{1/3}\frac{(m_eC_e^2+m_hC_h^2)}{h^2Mu^2} \qquad \ldots\ldots(7)$$

where ω is the volume of the unit cell, M the mass within it, and u the velocity of sound in the crystal. C_e and C_h are the interaction constants for electrons and holes respectively, and they are related to the mobilities μ by the expression

$$\mu_eC_e^2m_e^{5/2}=3eh^4Mu^2(512\omega^2k^3T^3\pi^7)^{-1/2}. \qquad \ldots\ldots(8)$$

Although equation (3) was derived for monatomic crystals, it is thought that it should hold fairly well for compounds of groups III and V as they resemble closely the corresponding group IV monatomic solids. For example the lattice constant of InSb is 6·48 while for grey Sn it is 6·49 Å.

(ii) A shift of the full and conduction bands relative to each other due to dilatation or contraction of the lattice. This has been treated by Shockley and Bardeen (1950) who find

$$(dE/dT)_2=2\beta(\pm C_e\pm C_h) \qquad \ldots\ldots(9)$$

where β is the linear expansion coefficient.

(iii) The *spreading* of the edge on heating may be due to some phenomenon such as that discussed by Cheeseman (1952) whereby normally forbidden transitions become progressively more probable as the lattice vibrations increase. As calculations of the effect involve calculations of transition probabilities including matrix elements, it would require considerable analysis to estimate the magnitude of the spread to be expected from this process.

From equation (8) C_e can be found in terms of m_e, provided that a value can be obtained for u, the velocity of sound in the crystal. Now the expression Mu^2/ω is the compressibility of the material, c_{11}. Study of the values of c_{11} for silicon, germanium and tellurium shows that this parameter does not vary greatly, and as the mechanical properties of InSb place it between germanium and tellurium it is relatively certain that, for InSb, c_{11} is between 1·5 and 0·5 (or say 0·9) $\times10^{12}$ cgs units (Shockley and Bardeen 1950). As this parameter appears as a square root in the expression for C_e or C_h, then the accuracy of the latter is probably within $\pm20\%$ which is quite adequate for the present treatment.

Putting the relevant values in equation (8), with $\mu_e=42000$, we find $C_em_e^{5/4}=0{\cdot}56m^{5/4}$ and with $\mu_h=600$, $C_hm_h^{5/4}=4{\cdot}74m^{5/4}$. Substituting these values of C_e and C_h in equation (7) we have a theoretical value for

$$(dE/dT)_1=-10^{-6}[0{\cdot}32(m/m_e)^{3/2}+19{\cdot}5(m/m_h)^{3/2}]. \qquad \ldots\ldots(10)$$

From equation (9) we can obtain $(dE/dT)_2$ from C_e and C_h if we know the coefficient of linear expansion, β. Again this will almost certainly lie between the values for Ge and Te—i.e. between 6×10^{-6} and 17×10^{-6}†, so we will take it as 10^{-5}.

Hence $$(dE/dT)_2 = 2\times10^{-5}[\pm0{\cdot}56(m/m_e)^{5/4} \pm 4{\cdot}74(m/m_h)^{5/4}] \quad \ldots\ldots(11)$$

where the signs to be used depend on the direction in which the full and conduction bands move on compression.

Equating $(dE/dT)_1+(dE/dT)_2$ to the experimental shift of -4×10^{-4} ev/°C we obtain:

$$400 = \pm11{\cdot}2(m/m_e)^{5/4} \pm 95(m/m_h)^{5/4} + 0{\cdot}32(m/m_e)^{3/2} + 19(m/m_h)^{3/2}.$$

Using the value of the mass product given by Tanenbaum and Maita of $(m_e m_h)^{1/2} = 0{\cdot}083m$ to eliminate m_e we obtain

$$4 = \pm56(m_h/m)^{5/4} \pm 0{\cdot}95(m/m_h)^{5/4} + 1{\cdot}6(m_h/m)^{3/2} + 0{\cdot}19(m/m_h)^{3/2}. \quad \ldots\ldots(12)$$

Clearly the first two terms of this equation are the important ones, and it is necessary to decide which signs to use. One of them of course must be positive, but they cannot both be positive or the right-hand side of the equation would have a minimum value many times the experimental figure on the left-hand side, so that apparently the two bands move in the *same* direction on compression. It transpires that there is little difference in the solution of the equation whichever term is taken as positive, the values obtained in the two cases being $m_h = 0{\cdot}22m$ or $0{\cdot}19m$ and $m_e = 0{\cdot}032m$ and $0{\cdot}037m$. We may therefore average the values and say $m_e = 0{\cdot}034m$ and $m_h = 0{\cdot}21m$.

It is difficult to assess the accuracy of this rather involved method of calculating the effective masses, but it is probably significant that it gives a value for m_e close to that calculated by the effective mass theory of §3.1. Using the above values of m_e and m_h we find that $C_e = 38$ and $C_h = 33$, so that as suggested earlier the high mobility of the electrons results from the very low effective mass, rather than from an abnormal value of C_e. We also find from these figures that the shift due to lattice broadening is $(dE/dT)_1 = -0{\cdot}3\times10^{-4}$ ev/°C and that due to dilatation is $(dE/dT)_2 = -3{\cdot}7\times10^{-4}$ ev/°C.

§4. Long Wavelength Absorption

The Drude theory of absorption by free electrons leads to an absorption coefficient

$$K = (4\pi\sigma/cn)(1+\nu^2/\gamma^2)^{-1} \quad \ldots\ldots(13)$$

(see Moss 1952), where n is the refractive index, σ the low frequency conductivity and γ the damping factor given by $2\pi\gamma = e/\mu m_e$.

For wavelengths appreciably greater than that of the absorption edge the experimental data for n-type InSb obey the quadratic law with frequency (or wavelength) given by equation (13). All the parameters in the equation except γ are measurable, so that a value of γ and hence m_e may be obtained.

Rewriting equation (13) and neglecting the unity compared with ν^2/γ^2 at the long wavelengths used we find

$$\frac{m_e}{m}\frac{\mu}{\lambda}\left(\frac{K}{\sigma}\right)^{1/2} = 18n^{-1/2} = 9{\cdot}4 \quad \text{since } n = 3{\cdot}7,$$

† Average value if assumed isotropic.

where μ is in $cm^2 V^{-1} sec^{-1}$, λ in microns, K in cm^{-1} and σ in $\Omega^{-1} cm^{-1}$. Substituting the relevant values of σ and μ_e for the two specimens described by Avery *et al.* we find $m_e = 0{\cdot}038m$ and $0{\cdot}027m$, or an average value of $m_e = 0{\cdot}032m$.

The reason why the p-type material shows no increase of absorption coefficient as λ^2 at long wavelengths is not understood at present. Use of equation (13) with the values of the effective masses determined earlier indicates that for equal numbers of carriers the p-type material should have rather more free carrier absorption than the n-type at a given long wavelength.

§ 5. Photoelectric Effects

Study of the spectral sensitivity curves given by Avery *et al.* shows an appreciable difference in the $\lambda_{1/2}$ values for the photoconductor and photodiode. This discrepancy is attributed to the different penetration depths of radiation which are effective in the two cases, and to the fact that the absorption edge is not infinitely steep.

For photoconductivity, active absorption anywhere in the body of the material will produce carriers which contribute equally to the photocurrent. The $\lambda_{1/2}$ point is thus determined by the condition that half the maximum absorption takes place. The latter is $1-R$, where R is the reflectivity, and the general expression for absorption is $(1-R)(1-e^{-Kt})/(1-Re^{-Kt})$. Hence at the $\lambda_{1/2}$ point we would expect the absorption constant to be given by $(1-e^{-Kt})/(1-Re^{-Kt}) = 1/2$. Putting $R = 33\%$ and $t = 0{\cdot}018$ cm we calculate $K = 30\,cm^{-1}$. Reference to figure 4 of their paper shows that at $\lambda_{1/2} = 7{\cdot}75$, $K = 27\,cm^{-1}$, so that the agreement is good.

For the photovoltage, however, the carriers must be absorbed near the probe or their probability of influencing the probe is small. For this case it would be expected that the reciprocal of the absorption constant at $\lambda_{1/2}$ would roughly equal the diffusion length for the carriers, i.e. $(D\tau)^{1/2}$ where D is the diffusion constant and τ the lifetime. From absorption data $K = 200\,cm^{-1}$ at the $\lambda_{1/2}$ point of $7{\cdot}15\mu$. Hence, $(D\tau)^{1/2} \sim 50\mu$. Now $D = \boldsymbol{k}T\mu/e$, so that putting the mobility $\mu = 42\,000\,cm^2 V^{-1} sec^{-1}$ as is appropriate for electrons (*N.B.* the holes will give a negligible contribution by virtue of their much lower mobility) we calculate $\tau \sim 1/40\,\mu$sec at room temperature.

The lifetime measured by direct observation rise of photoconductivity at $-183°$C was $\sim 0{\cdot}8\,\mu$sec, so that allowing for the fact that for the specimen measured the carrier concentration at room temperature was some 20 times greater than at $-183°$C, the above figure of $1/40\,\mu$sec for the room temperature lifetime seems reasonable.

With this value of τ it is possible to calculate the absolute level of the photoconductive signal to be expected for a given irradiation, and to compare it with an experiment. It may be shown that for a photoconductor with matched load and a voltage V applied to the specimen†, the expected r.m.s. photoconductive signal is $\Delta v = (V\Delta N/4{\cdot}4Ant)$, where $\Delta N/Ant$ is the relative increase in numbers of carriers in the specimen on illumination, and t and A are the specimen thickness and area. Putting $N = \eta Q\tau$, where Q is the rate of absorption of quanta in the sample, and the quantum efficiency $\eta = 1$ (again considering only the electrons) we have $\Delta v = (VQ\tau/4{\cdot}4Ant)$.

† For a transformer load V approximates to the total applied voltage.

In the experiment described by Avery *et al.*, Q was $1{\cdot}3 \times 10^{15}$ quanta $cm^{-2}\,sec^{-1}$, effectively, and V was 0·3 volt. Hence the calculated signal is $\Delta v = 12 \times 10^{-9}$ volt, as compared with the observed signal of $9{\cdot}6 \times 10^{-9}$ volt.

The agreement is thus satisfactory, and so confirms the general applicability of simple photoconductive theory to this system, and indicates that the estimation of τ and the assumption of unit quantum efficiency were substantially correct.

§ 6. Conclusions

The extremely high mobility of electrons in InSb, and the phenomenon of the marked dependence of the absorption edge on purity for n-type samples are both attributed to an abnormally small effective electron mass. This mass has been estimated in three ways (i) from analysis of data on the temperature dependence of the absorption edge, (ii) from a theoretical relation between the effective mass and the curvature of the energy surface of the conduction band in momentum space, and hence with the variation in activation energy with impurity concentration, and (iii) from the magnitude of the free carrier absorption. The three estimates so obtained show agreement which is surprisingly good in view of the tentative nature of some of the theories used, the values being $m_e/m = 0{\cdot}034$, 0·028 and 0·032.

It is therefore concluded that the effective mass of the conduction electrons in InSb is near 0·03 of the free electron mass, and that consequently $m_h/m = 0{\cdot}23$.

The magnitude of the photoconductivity observed in intrinsic InSb at room temperature can be explained by simple photoconductive theory with a quantum efficiency of approximate unity.

Acknowledgments

Thanks are due to the Chief Scientist, Ministry of Supply for permission to publish this paper, and to Dr. G. G. Macfarlane and Dr. L. Pincherle of the Radar Research Establishment, Malvern, for their constructive suggestions.

References

Avery, D. G., Goodwin, D. W., Lawson, W. D., and Moss, T. S., 1954, *Proc. Phys. Soc.* B, **67**, 761.

Burstein, E., 1954, *Phys. Rev.*, **93**, 632.

Cheeseman, I. C., 1952, *Proc. Phys. Soc.* A, **65**, 25.

Fan, H. Y., 1951, *Phys. Rev.*, **82**, 900.

Moss, T. S., 1952, *Photoconductivity in the Elements* (London : Butterworths).

Pearson, G. L., and Tanenbaum, M., 1953, *Phys. Rev.*, **90**, 153.

Shockley, W., and Bardeen, J., 1950, *Phys. Rev.*, **77**, 407.

Tanenbaum, M., and Briggs, H. B., 1953, *Phys. Rev.*, **91**, 1561.

Tanenbaum, M., and Maita, J. P., 1953, *Phys. Rev.*, **91**, 1009.

The Vapour Pressure of Calcium : I

By P. E. DOUGLAS†
Physics Department, The University, Reading

Communicated by R. W. Ditchburn; *MS. received 9th July* 1954

Abstract. The vapour pressure of calcium in the temperature range 800°K–920°K has been measured by the Knudsen effusion method. Calcium metal was heated in a specially designed nickel crucible, and the vapour which effused through a narrow slit aperture into a well-defined solid angle was condensed and determined chemically. The vapour pressure p in mm Hg was found to be related to the absolute temperature T by the relation $\log_{10} p = 9{\cdot}59 - 10089/T$.

§1. Introduction

Measurements of the vapour pressure of solid calcium in the range 500°C to 650°C have been reported by Pilling (1921) who derived the vapour pressure from measurements of the rate of evaporation from a heated wire, and by Rudberg (1934) who used an effusion method. Since Rudberg's values are lower than Pilling's by a factor of 10, the present work was carried out to provide further experimental data.

The crucible used by Rudberg was not satisfactorily designed: it consisted of a cylindrical pit in a molybdenum cylinder, and the method by which it was heated did not ensure the optimum temperature distribution. In the experiments to be reported in this paper particular care was taken in the design of the crucible, and the methods of heating and temperature measurement.

§2. Theory of the Method

The calculation of the vapour pressure p inside the crucible, at absolute temperature T, can be made from the measured rate dn/dt at which vapour atoms of mass m effuse through an orifice of area ΔS into a solid angle $\Delta\Omega$ at a direction θ to the normal. According to the well-known laws of gas flow we have

$$\frac{dn}{dt} = \frac{\Delta S \Delta\Omega \cos\theta}{2\pi} \frac{p}{kT} \left(\frac{2kT}{\pi m}\right)^{1/2}, \qquad \ldots\ldots(1)$$

where k is Boltzmann's constant.

If the effusion rate is written dW/dt g sec^{-1}, equation (1) becomes

$$p = \frac{dW}{dt} \frac{\pi}{M \Delta S \Delta\Omega \cos\theta} (2\pi M R T)^{1/2}, \qquad \ldots\ldots(2)$$

where M denotes the atomic weight, and R the gas constant per mole.

This equation applies only for monatomic vapours, and is valid only if at least one linear dimension of the orifice is small compared with the mean free path. In order that a conveniently measurable effusion rate may be obtained, the area of the slit must not be too small, optimum conditions are therefore achieved by using a rectangular slit orifice whose width is small compared with the mean free path. A further limitation on the area of the orifice is imposed by the condition

† Now at Standard Telephones & Cables, Ltd., Ilminster, Somerset.

that the rate of evaporation within the crucible must be sufficient to maintain equilibrium between solid and vapour in spite of the loss of vapour through the orifice. This loss will be equal to $K\Delta S$, and the rate of evaporation to $K(1-\rho)A$, where K is a constant depending on the thermal velocity, A the surface area of the solid, and ρ the reflection coefficient for vapour atoms striking the solid surface. There is some evidence that $1-\rho$ is as low as 0·1 for calcium, and to ensure correct equilibrium conditions it is advisable to make A equal to or greater than $100\Delta S$.

The equilibrium temperature will be that of the coolest region of the crucible, and the design of the oven must be such that a well-defined cool region exists, to which the measured temperature refers.

§3. Apparatus and Method

The vacuum system consisted of a bell jar sealed to a brass plate with Apiezon 'Q' compound, and was pumped with a silicone oil diffusion pump. The arrangement of the apparatus within the vacuum system is shown in figure 1. The crucible O contained the calcium metal in the holes H (figure 2), and had a narrow slit

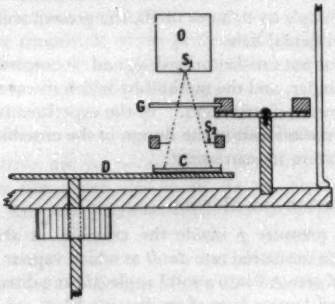


Figure 1.

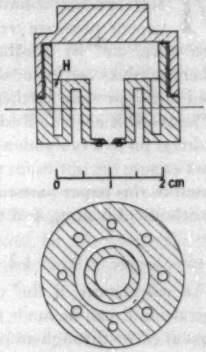


Figure 2.

orifice S_1 through which the vapour effused. A second slit S_2 placed immediately below and parallel to S_1 served to define a vapour beam of known dimensions from which the solid angle $\Delta\Omega$ was calculated. In this position the angle θ was zero for the centre of the beam, but a small correction of about 2% was required to allow for θ being finite near the ends of S_2.

The vapour passing through S_2 was condensed on a microscope coverslip C, eight of which were placed on a brass disc D which could be rotated through a Wilson seal. A shutter G, magnetically operated from outside the bell jar was used to control the exposure time.

The design of the crucible is shown in figure 2, it was machined from a nickel rod, and its overall dimensions were diameter 1·063 in., height 1·0 in. Two thin razor blades clamped to the lower face defined the orifice S_1. A heating element of nichrome strip wound on a mica former surrounded the whole crucible, and a similar smaller element was placed in the annular slot around the part containing

the orifice. This ensured that the orifice was the warmest region of the crucible, a condition necessary to prevent local condensation of vapour. Exploratory tests with thermocouples showed that the coolest region was near the centre of the upper face of the crucible, and in the vapour-pressure determinations the temperature was measured at this point and near the orifice with chromel–alumel thermocouples calibrated at the melting points of lead (327°c), zinc (419°c), antimony (630·5°c) and sodium chloride (800°c).

Before being placed in the crucible the calcium metal was cleaned in dilute HCl and alcohol, and finally washed with ether. When the vacuum system had been evacuated to a pressure of less than 10^{-4} mm Hg a coverslip was brought into position below S_2 and, with the shutter closed, the crucible was heated to the required temperature. The shutter was then opened for a known exposure time, sufficient to give between 100 and 500 μg of deposited calcium. Thirteen runs at different temperatures between 800°K and 920°K were carried out, in some of which the slit width of S_1, and the distance between S_1 and S_2 were varied slightly.

When exposed cover slips were removed from the vacuum system, the calcium deposit was dissolved in about 5 ml of hot dilute HCl in a conical ended centrifuge tube, and calcium oxalate was precipitated by the addition of 1 ml of a 4% solution of ammonium oxalate. After centrifuging at 1500 rev/min for 5 minutes the clear liquid was decanted off and the tube allowed to drain; 3 ml of dilute ammonia solution were added so that the entire inside surface of the tube was washed. After re-centrifuging, the oxalate was dissolved in 2 ml of normal sulphuric acid, the solution was then heated for a minute in a boiling water bath and titrated against a standardized solution of potassium permanganate in a 5 ml burette graduated to 0·01 ml.

§4. Results and Discussion

The dimensions of the slits S_1 and S_2 in three typical runs are given in table 1, and the results of the vapour-pressure determinations in table 2. Variation of the area of S_1 from 0·013 to 0·038 cm² appeared to have no significant effect on the measured values of the vapour pressure. All the results fitted a relation of the form $\log_{10} p = A - B/T$ to within the experimental errors, and evaluation of the constants A and B by the method of least squares gave the result

$$\log_{10} p = 9{\cdot}59 - 10089/T$$

for the vapour pressure of calcium in mm Hg in the temperature range 800°K to 920°K as shown in figure 3.

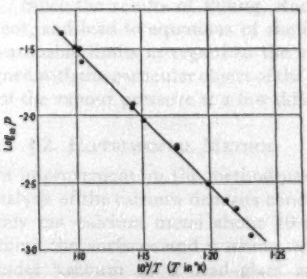


Figure 3.

Random experimental errors of about 3% in the determination of the calcium in the deposits, and about 1% in the measurement of the temperature may be expected, and these lead to the expectation of a total random error of about 5% in the values of the vapour pressure at a given temperature. It is hoped that serious systematic errors were avoided by the precautions taken in the design of the apparatus. The scatter of experimental points allows a variation of B between 11050 and 9580, and the corresponding values of A required to satisfy the experimental results at the centre of the temperature range are 10·7 and 9·0 respectively.

Table 1

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	918	0·046	0·47	0·0216	0·429	1·949	3·75
3	907	0·026	0·47	0·0127	0·429	1·949	4·07
12	820	0·082	0·47	0·0385	0·429	1·949	3·75

(1) Experiment number; (2) T(°K); (3) width of S_1(cm); (4) length of S_1(cm); (5) area of S_1(cm^2); (6) width of S_2(cm); (7) length of S_2(cm); (8) distance S_1S_2(cm).

Table 2

Expr. No.	T(°K)	p (mm Hg)	$\log_{10}p$	Expr. No.	T(°K)	p (mm Hg)	$\log_{10}p$
1	918	$4·09\times10^{-2}$	−1·388	8	851	$5·84\times10^{-3}$	−2·233
2	910	$3·28\times10^{-2}$	−1·484	9	851	$4·82\times10^{-3}$	−2·317
3	907	$3·13\times10^{-2}$	−1·504	10	835	$3·06\times10^{-3}$	−2·514
4	906	$2·55\times10^{-2}$	−1·593	11	825	$2·30\times10^{-3}$	−2·639
5	876	$1·13\times10^{-2}$	−1·947	12	820	$1·97\times10^{-3}$	−2·706
6	876	$1·24\times10^{-2}$	−1·907	13	807	$1·23\times10^{-3}$	−2·909
7	869	$9·39\times10^{-3}$	−2·029				

Acknowledgment

The author wishes to express his thanks to Professor R. W. Ditchburn for many helpful discussions.

References

Ditchburn, R. W., and Gilmour, J. C., 1941, *Rev. Mod. Phys.*, **13**, 310.
Pilling, N. B., 1921, *Phys. Rev.*, **18**, 366.
Rudberg, E., 1934, *Phys. Rev.*, **46**, 763.

The Vapour Pressure of Calcium: II

By D. H. TOMLIN
Physics Department, The University, Reading

Communicated by R. W. Ditchburn; MS. received 9th July 1954

Abstract. Further determinations of the vapour pressure of solid calcium have been carried out by a similar application of the Knudsen effusion method to that described in the companion paper by Douglas, but with radiochemical analysis of the condensed deposits. It is concluded from the good agreement with the results of Douglas and of Pilling that vapour pressure data for the temperature range 800°K to 900°K are now well established, and that the results of Rudberg are in error. The matching of the results with those of Hartmann and Schneider for liquid calcium when extrapolation is made to the melting point is discussed. Agreement in this respect remains incomplete.

§ 1. Introduction

The measurements of the vapour pressure of solid calcium reported in the companion paper by Douglas (1954) agree with the older results of Pilling (1921) more closely than with the later ones of Rudberg (1934). This is a little surprising, since Douglas and Rudberg both used Knudsen's effusion method, and Pilling used the less direct Langmuir evaporation method, involving large correction factors in the evaluation of results. When the vapour-pressure–temperature relations of each of these authors is extrapolated to the melting point of calcium, none agrees with the similarly extrapolated curves for liquid calcium published by Ruff and Hartmann (1924) and Hartmann and Schneider (1929).

In view of the diversity of these results, which has been pointed out previously by Ditchburn and Gilmour (1941), it was decided to carry out further determinations by a method similar to that used by Douglas, but with certain improvements, with the aim of establishing the data for solid calcium with greater certainty. Since the results of Pilling, Rudberg and Douglas are each internally consistent, and lead to equations of the form $\log_{10} p = A - B/T$, and all agree within reasonable limits in regard to the value of B, the present experiments were designed with the particular object of the accurate determination of the absolute values of the vapour pressure at a few different temperatures.

§ 2. Experimental Method

The most important improvement on the method used by Douglas was the use of radiochemical analysis of the calcium deposits condensed from the vapour beam. Pieces of freshly cut calcium metal about 10 mm × 1 mm × 1 mm in size were cleaned by filing the surfaces and washing in ether, and were then immediately sealed under vacuum in a lead-glass ampoule. They were irradiated with neutrons in the Harwell pile for four weeks, and were kept

in vacuo until the beginning of the experiments, which was about a month after the end of the irradiation. This time was allowed for the decay of the short-lived radioactive isotopes ^{41}Ca, ^{47}Ca and ^{49}Ca; the specific activity due to ^{45}Ca (half-life 160 days) was then approximately $200\,\mu c\ g^{-1}$.

Figure 1 shows the arrangement of the effusion apparatus used. The oven was of the same design as that used by Douglas, but was machined from a low-carbon iron instead of nickel in view of evidence that this material was less reactive with calcium. A circular orifice was made, as suggested by Schadel and Birchenall (1950), by dimpling a 0·003 in. thick nickel foil with a sharp punch, and polishing away the convex side until a hole of the required diameter was formed. Examination under a microscope showed that holes made in this way had sharply defined edges, and were very nearly circular. The microscope was used to measure the average diameter, and the foil was then clamped on to the lower face of the oven, with the orifice concentric with a larger hole in the oven wall. The methods of heating the oven and measuring the temperature were also similar to those used by Douglas.

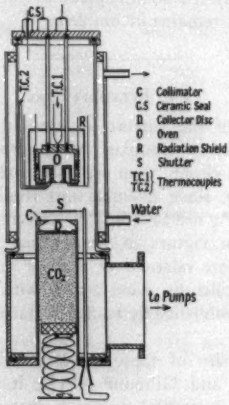


Figure 1. Effusion apparatus.

The oven was suspended rigidly above a combined collector disc holder and beam collimator, which formed the top of a re-entrant tube which could be cooled by packing with solid carbon dioxide in the manner shown in the diagram. A radiation shield of nickel foil surrounded the oven, and the enclosing glass vessel was cooled by means of a water jacket. The collector discs, which were 1 in. diameter glass cover slips, were shielded from the vapour beam by a nickel shutter which could be rotated, through a conical seal, to expose the discs during timed exposures.

Coaxial alignment of the oven orifice and collimating aperture was effected by sliding the flanged glass tube, forming the upper part of the vacuum system, on its rubber gasket. The whole apparatus was easily demountable to facilitate the removal and replacement of the cover slips after each run.

The oven charge consisted of eight pieces of the radioactive calcium, and the apparatus was evacuated to a pressure of 10^{-5} mm Hg, measured with an ionization gauge, by an oil diffusion pump. The following procedure was followed in each vapour-pressure determination.

With the shutter covering the collector disc, the oven was slowly heated to the required temperature by passing a constant current through the subsidiary orifice heater from a transformer, and an adjustable current through the main heater. When the temperature indicated by the thermocouple at the top of the oven had remained constant for about five minutes, the shutter was opened for the required exposure time, during which the temperature was maintained within $\pm 1°$ at a constant value by continuous observation and manual adjustment of a Variac. The temperature at the orifice was about 10° higher. Throughout all the exposures the gas pressure in the apparatus remained at or below 10^{-5} mm Hg. At the termination of the exposure by closing the shutter, the heaters were switched off, and when the oven had cooled the vacuum was broken by admitting nitrogen, in order to replace the collector disc for the next run.

A determination of the ^{45}Ca activity of each deposit was made by placing the disc in one of two standard positions beneath a thin-windowed Geiger counter. One of the positions was close to the window and was used for sources of low activity and the other, more distant, position for those of higher activity, sources of intermediate strength being measured in both positions to give the sensitivity ratio by means of which all activities could be related to one standard counting geometry. Frequent measurements of the activity of a standard uranium source were made to allow corrections to be applied for small variations in the overall sensitivity of the counting apparatus, and further corrections were applied when necessary for the loss of counts due to its known dead time of 300 μsec.

In order to convert these relative activity measurements to absolute amounts of calcium, a determination of the specific activity of the deposited calcium, defined as the number of counts per minute in the standard counting geometry, per mg of calcium in the deposit, was made as follows: On completion of all the vapour-pressure runs the oven orifice was enlarged *in situ* so that in long exposures, without careful temperature control, sufficient calcium could be collected to enable accurate chemical analysis to be carried out. These deposits were conveniently measured for activity using the more distant of the source positions in the counting apparatus, and were subsequently dissolved off the discs with dilute HCl and transferred to centrifuge tubes ready for chemical determination. The discs were re-counted afterwards to ensure that all activity had been removed. An 'artificial' deposit made by evaporating a suspension in water of $CaCO_3$ prepared from the oven charge was similarly counted and chemically analysed.

§ 3. Results

Table 1 shows the results of specific activity determinations on samples obtained by effusion from the oven and by evaporation from suspension. The result derived from the latter must be expected to be higher than the others on account of the isotopic fractionation which occurs during effusion. This will have the effect of reducing the ^{45}Ca content of the condensed calcium compared with that of the oven charge, and in order to compare the third result in the table with the other two it must be multiplied by the factor $(40/45)^{1/2}$. As shown

at the foot of the table, good agreement is then obtained among all three determinations, and the average value may be taken as the specific activity required for calculating the vapour pressures.

Table 1. Results of Specific Activity Determinations

Sample	Condensed from effused vapour 1	2	Deposited from suspension 3
Total Ca (mg)	$0{\cdot}97_2$	$0{\cdot}96_4$	$0{\cdot}97_2$
Activity (counts/min in standard geometry)	7448	7200	7776
Specific activity (counts/min per mg)	7663	7469	7991
Correction of result 3 to isotopic composition of effused vapour			$7991\times(40/45)^{1/2}=7510$
Average specific activity			7544

In the vapour pressure determinations the area of the oven orifice was 0·00522 cm², and the solid angle subtended at the orifice by the collimating aperture was 0·0854 of 4π.

If α denotes the activity, related to the standard counting geometry, of a deposit obtained in a particular run of duration t sec at a temperature T°K, the corresponding vapour pressure may be calculated from Douglas' (1954) equation (2), which, on substituting the above numerical values, including the specific activity, and converting the units, becomes

$$p=2{\cdot}546\times10^{-3}\alpha T^{1/2}/t \text{ mm Hg.}$$

Table 2. Vapour Pressure Determinations

Run No.	T (°K)	$10^3/T$	t (sec)	α counts/min	$\frac{\alpha\sqrt{T}}{t}$	p (mm Hg)
1	868·9	1·151	1800	255	$4{\cdot}17_6$	$1{\cdot}06_9\times10^{-2}$
2	864·4	1·156	1860	233	$3{\cdot}68_2$	$0{\cdot}93_5\times10^{-2}$
3	867·4	1·153	1800	241	$3{\cdot}94_2$	$1{\cdot}00_0\times10^{-2}$
4	800·7	1·248	3660	58·5	$0{\cdot}45_3$	$1{\cdot}15_0\times10^{-3}$
5	838·2	1·193	3600	184·5	$0{\cdot}48_4$	$3{\cdot}76_7\times10^{-3}$
6	877·2	1·140	1200	199·5	$4{\cdot}92_2$	$1{\cdot}25_0\times10^{-2}$

The results for the six determinations made are given in table 2, and are plotted in a graph of $\log_{10}p$ against $10^3/T$ in figure 2, together with the results of Pilling, Rudberg and Douglas. The straight line $\log_{10}p=9{\cdot}36-9866/T$ which has been drawn is compatible with all the results except Rudberg's, and, as will be shown later, leads to some degree of correlation with Hartmann and Schneider's results for liquid calcium.

The activity appearing on the underside of the collimator plate after an exposure was in no case greater than 0·5% of the corresponding collector disc activity, showing that reflection of atoms from the disc was probably negligible.

When the oven was opened on compietion of all the experiments the calcium was still quite clean and metallic in appearance and, furthermore, there was a circular patch of condensed metallic calcium covering most of the inside upper face of the oven, showing that this must have been the coolest region during the runs.

§ 4. Discussion

4.1. *The Experimental Data*

The present results clearly support those of Pilling and Douglas, and these three determinations provide a considerable body of evidence establishing the vapour pressure of solid calcium in the temperature range 800°K–900°K. It must be admitted that the close agreement of Pilling's results with those of Douglas and the present author is to some extent fortuitous in view of the indirect nature of his method. He compared the rates of evaporation from calcium, cadmium and zinc wires, and from the known vapour pressures of cadmium and zinc derived a correction factor for use in calculating the vapour pressure of calcium. This factor was of the order 100, and though dependent partly on geometrical arrangements, which were the same for all three elements, must have depended also on surface properties which may have been quite different. The errors thus introduced may well have been independent of temperature within the experimental range, and this would account for the agreement in respect of the gradient of the $(\log_{10} p,\ 1/T)$ curve.

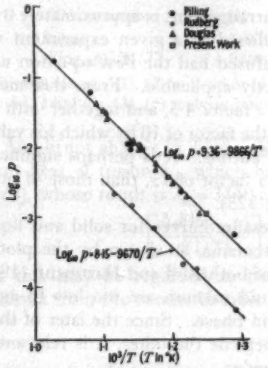


Figure 2. Experimental results for vapour pressure of solid Ca.

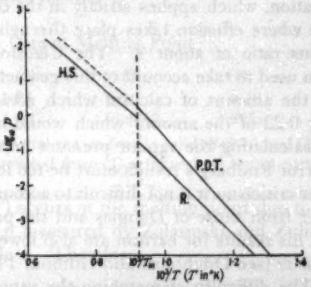


Figure 3. H. S. Hartmann and Schneider $\log_{10}p = 8{\cdot}01 - 8777/T$. R. Rudberg $\log_{10}p = 8{\cdot}15 - 9670/T$. P. D.T. Curve suggested in present paper, combining experimental results of Pilling, Douglas and Tomlin, and agreeing in gradient with the H.S. curve. The broken curve agrees with the P.D.T. curve in gradient and absolute values of the vapour pressure $\log_{10}p = 9{\cdot}36 - 9866/T$ below T_m, and $\log_{10}p = 8{\cdot}34 - 8777/T$ above T_m.

Several criticisms may be put forward to account for the lower vapour pressures obtained by Rudberg with his variation of the effusion method. Two important criteria which must be observed in using this method are: (*a*) The diameter d of the orifice must be small compared with the mean free path length λ of the vapour at the temperature of the experiment, and its area must be small compared with the surface area of the evaporating solid. Indeed, the area ratio necessary may well be of the order 100. (*b*) The measured temperature for which the vapour pressure is determined in any particular experiment must

be that of the coolest region of the oven. The warmest region should be in the vicinity of the orifice to prevent condensation of vapour restricting its effective area.

Failure to observe either of these criteria would lead to too low a value for the measured vapour pressure corresponding to a given measured temperature. Particular care in these matters was taken by Douglas and the present author; the value of λ for uncharged calcium atoms at 875°K, where the vapour pressure is approximately $1{\cdot}2 \times 10^{-3}$ mm Hg, is about 4·1 cm, so that with an orifice diameter $d = 0{\cdot}0815$ cm, as used in the present work, $\lambda/d \simeq 50$. In addition, the ratio (area of Ca metal in oven)/(area of orifice) was certainly greater than 200.

In Rudberg's apparatus the orifice area was equal to the total cross-sectional area of the oven interior, and must have been of the same order of magnitude as the surface area of the calcium. The ratio λ/d was approximately 20, and though smaller than the value applicable in the present work, was probably acceptable, since Schadel and Birchenall (1950) in measuring the vapour pressure of silver found no appreciable effects when λ/d was varied from 104 to 12·5. A rather serious error arises, however, from the use of the Knudsen gas flow equation, which applies strictly in the case of an aperture in a thin plate, to the case where effusion takes place through a cylindrical channel with a length-to-radius ratio of about 9. The Clausing correction factor which should have been used to take account of this geometrical arrangement is approximately 0·22, i.e. the amount of calcium which actually effused in a given experiment was only 0·22 of the amount which would have effused had the flow equation used for calculating the vapour pressure been strictly applicable. From this source of error Rudberg's results must be too low by a factor 4·5, and together with the other criticisms it is not difficult to account for the factor of 10 by which his values differ from those of Douglas and the present author. It is perhaps significant that his results for barium are also lower, by a factor of 15, than those of other workers (see Ditchburn and Gilmour 1941, p. 325).

The difficulty of matching the vapour pressure curves for solid and liquid calcium at the melting point (1083°K) still remains, as shown by the plot of experimental values in figure 3. The results of both Ruff and Hartmann (1924) and Hartmann and Schneider (1929) for liquid calcium are too low to agree with the now well-established data for the solid phase. Since the later of these two publications is no doubt intended to supersede the other, it is relevant to discuss only the work of Hartmann and Schneider.

It is possible to point out two sources of error which might lead to too low an estimation of the vapour pressure from their measurements of the boiling point in an iron crucible under known pressure of inert gas. Firstly, the thermocouple which was moved about in the vapour close to the liquid surface until a region of constant temperature was found was not shielded from heat radiation from the walls of the crucible. This would probably lead to the indication of too high a temperature for a given pressure. Secondly, it seems possible that the iron crucible was not completely chemically inert to the liquid calcium. Some reaction with an iron oven was noticed at the much lower temperatures used in the present work, and in this laboratory liquid calcium has been found to attack nickel, tantalum and other metals at temperatures lower than the range of Hartmann and Schneider's experiments (P. J. Jutsum, unpublished work).

4.2. *Extrapolation to the Melting Point*

It may be shown that the gradient of Hartmann and Schneider's ($\log_{10}p$, $1/T$) line for liquid calcium is in reasonably good agreement with that of the line suggested in figure 2 for the solid. The extrapolated values of the vapour pressure p and the gradient $d(\log_{10}p)/d(1/T)$ at the melting point may be derived from the well-known equation

$$\boldsymbol{R}T^2\,d(\ln p)/dT=\lambda_1-\lambda_m-\int_{T_1}^{T}(C_c-C_g)\,dT \qquad \ldots\ldots(1)$$

where λ_1 is the heat of vaporization at $T=T_1$, λ_m the heat of transition at any transition point in the temperature range T_1–T, and C_c, C_g are the atomic heats of the condensed and gaseous phases.

The following specific heat data are given by the U.S. Bureau of Mines (1949): $C_c=6{\cdot}29+1{\cdot}40\times10^{-3}T$ for β Ca (673°K–1083°K), $C_c=7{\cdot}50$ for liquid Ca, $C_g=4{\cdot}97$. With the aid of these data, and substituting experimental values of $\log_{10}p$ and $d(\log_{10}p)/d(1/T)$ corresponding to the mid-points of the solid and liquid experimental ranges, the corresponding values at the melting point $T_m=1083$°K have been calculated, as shown in table 3.

Table 3. Extrapolated Data at Melting Point

Temperature	(1)	(2)	(3)	(4)
$\log_{10}p$	−1·97	1·75	0·21	−0·10
$d(\log_{10}p)/d(1/T)$	−9866	−8777	−9743	−8952

Temperature: (1), (2) mid-point of experimental curves for (1) solid (869°K) and (2) liquid (1400°K); (3), (4) melting point extrapolated from (3) solid and (4) liquid curves.

An abrupt change in $d(\log_{10}p)/d(1/T)$ occurs at the melting point due to the latent heat of fusion λ_m, which has been measured by Zalesinski and Zulinski (1929), whose result is $\lambda_m=3140$ cal mole^{-1}. Therefore

$$\left[\frac{\Delta d(\log_{10}p)}{d(1/T)}\right]_{T=T_m}=-\frac{\lambda_m}{2{\cdot}303\boldsymbol{R}}=-688$$

and this may be regarded as being in fair agreement with the value −791 obtained from the results in table 3.

§ 5. Conclusion

From the close agreement of the results obtained with those of Pilling and Douglas it is concluded that the vapour pressure relation for solid calcium in the temperature range 800°K to 900°K is now well established, and it is suggested that the equation $\log_{10}p=9{\cdot}36-9866/T$ is the best representation of the experimental data up to the melting point. If this is accepted, Hartmann and Schneider's equation for liquid calcium should be modified to

$$\log_{10}p=8{\cdot}34-8777/T,$$

i.e. their vapour pressures should be multiplied by the factor 2·14.

Acknowledgments

The author wishes to thank Professor R. W. Ditchburn for helpful discussions, and Mr. T. G. Taylor (Department of Agricultural Chemistry, University of Reading) for carrying out the chemical determinations.

References

Ditchburn, R. W., and Gilmour, J. C., 1941, *Rev. Mod. Phys.*, **13**, 310.
Douglas, P. E., 1954, *Proc. Phys. Soc.* B, **67**, 783.
Hartmann, H., and Schneider, R., 1929, *Z. anorg. Chem.*, **180**, 275.
Pilling, N. B., 1921, *Phys. Rev.*, **18**, 362.
Rudberg, E., 1934, *Phys. Rev.*, **46**, 763.
Ruff, O., and Hartmann, H., 1924, *Z. anorg. Chem.*, **133**, 29.
Schadel, H. M., and Birchenall, C. E., 1950, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **188**, 1134.
U.S. Bureau of Mines, 1949, *Bull.* No. 476, pp. 30–32.
Zalesinski, E., and Zulinski, R., 1929, *Bull. Nat. Acad. Pol. (Krakau)* A, 479.

CORRIGENDUM

In the September issue of the Proceedings of the Physical Society, Section 'B', the title of the paper by Miss J. P. A. Tillett (p. 677) was printed incorrectly. The correct title is " A Study of the Impact of Spheres on Plates ".

REVIEWS OF BOOKS

Metallurgy of the Rarer Metals, edited by H. M. FINNISTON. (London: Butterworths Scientific Publications; New York: Academic Press).

No. 1. *Chromium*, by A. H. SULLY. Pp. xii + 272, 1954. 35*s*.

No. 2. *Zirconium*, by G. L. MILLER. Pp. xviii + 382, 1954. 45*s*.

These books form part of a series describing chromium, zirconium, titanium etc., which is being edited, under the title *Metallurgy of the Rarer Metals*, by Dr. H. M. Finniston of Harwell. The value of this project, already exemplified in the above two volumes, is that accumulated knowledge, inevitably very disconnected at an early stage of development and use, is much more useful when sorted and criticized by an authority on each individual metal.

The books are excellently produced, in substantially the same style, and contain numerous and well selected diagrams and graphs.

The bibliographies at the end of each chapter are thorough but the reviewer prefers a single list numbered and arranged alphabetically as to author at the end of the book, as this is easier to find and acts as an additional index to authors. The practice adopted by Miller of adding book-page references is excellent.

Chromium, by A. H. SULLY.

Chromium might perhaps seem out of place among the 'rarer' metals, for one's eye is struck by its common appearance in modern life, and over 2 million tons of its ore are mined each year. However, the *pure* metal is a comparatively recent product. It was, of course, hoped that with high purity the outstanding drawback of brittleness would disappear, but unfortunately this has not happened.

The occurrence of chromium in nature, its commercial extraction, and that of its ferro-alloys, by the electric furnace and aluminium reduction processes are discussed; an account of its preparation in highly pure form follows. The effect of heat treatment on the gaseous impurities left in electrolytic chromium is described.

In the third chapter the physical properties of chromium are surveyed. The existence of an hexagonal alternative to the normal body-centred cubic structure is unlikely according to the evidence the author marshals, nor does he think that an allotropic change just below the melting point is conclusively established. He points out that, although an anomaly exists at about 38°C in the thermal expansion, elasticity, resistance and thermoelectric power, no corresponding anomaly has yet been found in the lattice spacing, and he says that no satisfactory explanation of the phenomenon is known.

Besides allotting 30 pages to the discussion of 'chromizing', the protection of metals by local diffusion of chromium at the surface under heat, Dr. Sully wisely devotes a similar space to electroplating and the peculiarities of electro-deposits. He admits that the reason for the low Young's modulus, high electrical resistivity and high hardness of the unannealed deposit is still far from clear, and that the part played by the network of tiny cracks has not been analysed. He agrees, however, with Brenner, Burkhead and Jennings that the cause of the change of

resistivity during annealing of the deposit is the agglomeration of the oxide, originally dispersed and hydrated, and considers that this is also of importance in explaining hardness changes on annealing as an additional cause to recrystallization and grain growth. In this he is somewhat at variance with Snaveley and Faust.

After discussing the melting of chromium and the difficulties of working it, the brittle–ductile transition, again of unknown cause, is examined. This has never been found lower than about 50°c and, unlike the 38°c anomaly, is *raised* by additives.

The final chapter is on chromium alloys, the low content alloys such as austenitic ' stainless ' steel being well known and of great importance, while the ' Nimonic ' alloys containing about 18% chromium are also very important for their creep resistance as gas-turbine rotor blades. However, Dr. Sully, in a well-reasoned conclusion, is pessimistic as to the future prospects for rich chromium alloys.

The book serves a useful purpose by indicating the dangers of work on samples of a few percent impurity or porosity, for the variations of available data on chromium are disconcerting: Young's modulus 40%, compressibility 50%, resistance × 5 times, melting point 2½%, density 4% (though up to 15% on some sintered materials) and hardness numbers 48 to 650. No values at all have been hazarded on Poisson's ratio, modulus of rigidity and on tensile strength. In his caution on the latter for brittle materials the author is in good company. The book is very free from obvious mistakes or printing errors. It should be noted, however, that the ordinate on Figure 29 is *apparent* Young's modulus (for it includes the effect of lateral inertia in the vibration test) and the caption of Figure 30 reads oddly. Reference 4 on p. 107 should be 1948 not 1938.

The very full bibliographies, which alone make the book essential for workers on chromium, are evidence of painstaking study by the author. This thorough study, aided by his own by no means inconsiderable contributions in the field, has enabled him to present a balanced account and to pick his way with discrimination among vague and discordant data besides fitting him to present his own case where, as indicated above, he differs from other authorities. The volume forms a readable and excellent book of reference on this subject.

Zirconium, by G. L. Miller.

This metal can legitimately be regarded as rare, for its ore is mined to the extent of only 1% of that for chromium. Most physicists will have only a vague idea of its properties; for instance, that it can be used in ' flint ' lighters, as a ' getter' in valves to improve their vacuum, as a secondary grid emission inhibitor again in valves, and perhaps that zirconium is a valuable refractory. It will be less well known that it is strong and corrosion-resisting and that, unlike chromium, it has the high ductility normally associated with metals. Its low absorption of slow neutrons appears to be of importance for pile applications.

After discussing its occurrence and extraction, the two main methods for large-scale production are described. These are: firstly, the van Arkel method, depending on the thermal production and decomposition of the iodide on a hot ' hairpin ', and secondly the Knoll process, using magnesium to decompose zirconium tetrachloride. Physicists will be shocked by the repeated use of the term ' crystal bar ' to ' distinguish ' the former from the latter product; the term ' iodide bar ' is often used instead.

In next discussing physical and mechanical properties the author is hampered by the inadequacy of data. Zirconium is a close-packed hexagonal metal with c/a ratio (1·59) only a little less than that of magnesium and almost identical with that of titanium. Unfortunately, apart from thermal expansion, no data on single crystal properties seem available. For this crystalline system anisotropy is inevitable in elasticity, resistivity and thermal expansion and conductivity, and the strong tendency of the metal to preferred orientation when worked leads to serious divergencies in values for these properties. The fairly full information given later (pp. 326–327) on the textures of rolled strip and drawn wire could perhaps have been profitably used to explain some of the divergencies in reported values on pp. 120 and 130. It should perhaps have been pointed out that the Bridgman's compressibility ($1/k$) data are seriously in conflict with Reynolds' elasticity data for $(k+\frac{4}{3}G)$ and G, G being the shear modulus. The unqualified use of the term ' longitudinal wave velocity ' for $\{(k+\frac{4}{3}G)/\rho\}^{1/2}$ is to be deprecated, for this term is often used for the ' thin bar ' velocity :

$$\left\{\rho\left(\frac{1}{9k}+\frac{1}{3G}\right)\right\}^{-1/2}.$$

The next chapter on strength deals with nearly 30 alloy systems and studies both hot and cold strengths. As an example, a zirconium–titanium alloy can have an ultimate tensile strength of rather over 0·15 million lb in^{-2}, with an elongation of a few per cent. The creep resistance is only discussed very briefly but appears promising at 500°c.

Two chapters are devoted to chemical attack, numerous agents being studied. Although oxygen, nitrogen and hydrogen attack zirconium readily at temperatures of a few hundred degrees centigrade, it is very resistant at room temperature; the same applies also to corrosion by most liquids. Alloying, as discussed in the next chapter, is often of chemical origin for a strongly electropositive metal like zirconium. Its high valency also favours alloying in spite of misfit. Dr. Miller's analysis of Brillouin zone studies discloses little of quantitative or even qualitative value, but even here circumstances appear to be encouraging for solubility. It is not therefore surprising that he is able to survey numerous alloys and present 17 binary phase diagrams, varying from the very complex Zr–Al system to the extremely simple Zr–Ti system.

The author's chapter on melting practice is worthy of study not only metallurgically but also for the diagrams and descriptions of furnace design which it contains. The arc melting process is one way of avoiding contamination, always a difficulty with such an active metal; but more ingenious is the ' levitation ' or float melting process he describes for melting and suspending free from all supports (although admittedly this is not yet applicable in zirconium).

The chapter on fabrication deals with operations on ingots up to 500 lb in weight. For the smaller ingots the process of hot rolling with the zirconium enclosed in a steel sheath is advantageous to avoid gaseous attack ; rolling for short periods at 650°c unsheathed can be employed but leaves a tenacious oxide scale. It would have been worth while to have included typical rolling and recrystallization textures in the form of pole figures; for undoubtedly appreciation of the importance of precise texture data is increasing. The extrusion of zirconium alloys is discussed (p. 308) and the favourable properties of this form are mentioned on pp. 134–135. Unfortunately no data on texture of extruded zirconium seem available.

The author expresses his view in the chapter on powder metallurgy that this method of preparation will not be widespread but will be valuable for special applications. He points out the fire and explosion hazard with powdered zirconium; fire risk when machining has already been mentioned.

The final chapter on special compounds of zirconium includes such materials as the carbide, boride and silicide; zirconium carbide is apparently harder than tungsten carbide.

Dr. Miller's own contributions to manufacturing methods for zirconium enhance the value of his survey and the volume is thoroughly to be recommended as a sound book of reference on the subject. G. BRADFIELD.

Electricity and Magnetism for Degree Students, by S. G. STARLING. 8th Edn. Revised in collaboration with the author by A. J. WOODALL. Pp. viii + 650. (London: Longmans, Green & Co., 1953.) 30*s*.

It is now over 40 years since the first edition of this book appeared. There can hardly have been another textbook in the whole range of physics which has been in such constant use in all this time. Students of today use it as their predecessors did. It owes it success to its insistence on the fundamental principles of the subjects of electricity and magnetism. The original author is now joined by a younger colleague as a collaborator and both continue Starling's original idea of keeping the experimental aspect of the subject in the foreground, but they stress the need to express it with the aid of mathematics, not however with too difficult a use of that branch of knowledge.

The content of this work is well known. It has been found of great use by students in all universities, colleges and schools in their degree courses, including Honours courses. It makes no claim to specialization in any part, but supplies a firm basis for those who may later become specialists. The subject has grown in the past forty years beyond all expectation, but the author has not allowed himself to be turned aside from his original purpose in order to give a disproportionate space to a popular new discovery. The task of pruning and grafting has not been easy, and in the present edition the difficulty of choice and extent of new additions must have been great. Opinions on this question are bound to differ amongst teachers and students, but on the whole no serious adverse criticism will be made on this account, especially as addition is not possible without making the book cumbersome as a textbook for students.

A number of more recent topics have been touched upon, and readers are directed to fuller accounts by means of references.

There are some points concerning notation and presentation which may be considered with advantage in preparing another edition. In applying Kirchhoff's laws in networks it would be an improvement to introduce the method of Maxwell's circular currents. The old notation which uses (P, Q, R) and (α, β, γ) for intensity components in Maxwell's equations should now be replaced by the more usual E's and H's. The general use of complex quantities in alternating current theory would shorten the work and lighten the burden of the reader. It will also prepare him for the general method adopted in all more advanced work. The short paragraph on the M.K.S. system in the chapter on units and dimensions could with advantage be expanded. Finally, the quadrant electrometer needs a more realistic treatment. H. T. FLINT.

Physics for Medical Students, 3rd Edn, by J. S. ROGERS. Pp. xiii + 405. (Melbourne: University Press, 1953; London: Cambridge University Press, 1954.) 35*s*.

This book is quite a mine of information on a large range of topics in medical physics. It was first reviewed in the *Proceedings* in 1934 (**46**, 301), and the present edition contains additional material, mainly on electronics and nuclear physics. The treatment of nuclear fission and nuclear reactors is extremely clear and well done, but the discussion on thermionic valves remains somewhat conventional and restricted.

As explained in the preface, the book is not itself a course in physics. It is an account of selected parts of the subject and it requires to be supplemented by an ordinary textbook of physics. Quite a high standard of knowledge is necessary if many of the formulae and arguments are to be understood.

It is not easy to decide just what should be included in a medical physics course, and opinions necessarily differ. It seems to the reviewer that Rogers directs rather too much attention to the intricacies of nuclear physics and cosmic rays. It is true that the use of radioactive isotopes in clinical medicine is increasing rapidly, but the clinician is more likely to fall into error through lack of understanding of his detecting and counting equipment than through ignorance of the length of the major semi-axis of an alpha-particle.

Improvements in medicine can come from better instrumentation and from better knowledge of the possibilities and shortcomings of instruments. In recent years there has been quite a transformation of the equipment in physiological laboratories, as well as in the clinical medicine, physical medicine and pathological departments of hospitals. Some of the impact causing these changes arose from the widespread war-time contact with radar and other devices. It would seem to lie with physics to keep up the impetus in this direction. Rogers' book supplies adequate material to indicate to the medical student the many tools which physics provides for his use, and will, it is hoped, encourage him to use them when the opportunity arises.

WILLIAM COCHRANE.

Contributions a l'étude de la déformation plastique, by B. JAOUL. Pp. iii + 86. (Paris: Publications Scientifiques et Téchniques du Ministère de l'Air, No. 290, 1954.) 800 fr.

Characteristics and Applications of Resistance Strain Gauges (*Proceedings of the N.B.S. Semicentennial Symposium on Resistance Strain Gauges, held at N.B.S. on 8th and 9th November* 1951.) Pp. iv + 140. (Washington: United States Department of Commerce, National Bureau of Standards Circular 528, 1954.) $1.50.

Proceedings of the Ankara Symposium on Arid Zone Hydrology. Pp. 268. (Paris: UNESCO; London: Ministry of Education, 1954.) 30*s*.

The Scientific Journal of the Royal College of Science, Vol. 23. Pp. iv + 114. (London: Royal College of Science, 1953.)

CONTENTS OF SECTION A

PAGE

Prof. A. KASTLER. Les Méthodes Optiques d'Orientation Atomique et leurs Applications 853

Mr. W. R. S. GARTON. Investigations of Atomic and Molecular Absorption Spectra: III—Ultra-Violet Absorption Spectra of Indium Vapour . . 864

Dr. A. A. WARE. Galvomagnetic and Thermomagnetic Effects in a Plasma . 869

Dr. J. C. KNIGHT, Dr. T. H. BRAID and Dr. H. O. W. RICHARDSON. The Beta Rays of Thallium 204 881

Dr. R. J. BLIN-STOYLE and Mr. M. A. PERKS. The Deviations of Nuclear Magnetic Moments from the Schmidt Lines 885

Prof. E. A. OWEN and Dr. G. I. WILLIAMS. X-Ray Measurements on Lithium at Low Temperatures 895

Prof. H. N. V. TEMPERLEY. On Feynman's Theory of Liquid Helium . . 901

Mr. T. R. CARSON. The Vibrational and Rotational Excitation of Molecular Hydrogen by Electron Impact 909

Dr. A. L. STEWART. The Photo-Ionization Cross Section of Lithium . . 917

Dr. E. W. T. RICHARDS. Radiations from Water under Alpha-Particle Bombardment 922

Dr. M. J. SEATON. The Photo-Ionization of Neon 927

Dr. G. GOLDRING. Emission of Electron–Positron Pairs from Light Nuclei: III—γ-Transitions in the Reaction $^{15}N(p, \gamma)^{16}O$ 930

Dr. C. B. SHARMA. The Absorption Spectrum of Bismuth Selenide and Bismuth Telluride in the Region 2900–2200 Å 935

Research Notes:

Dr. P. A. FRASER. Vibrational Transition Probabilities of Diatomic Molecules: III 939

Dr. P. J. CAMPION and Dr. W. T. DAVIES. The Detection of Electron Pairs in a Cloud Chamber with Internal Counters 941

Dr. P. J. CAMPION and Dr. W. T. DAVIES. The Resolving Time of an Internal Counter Controlled Wilson Chamber 944

Mr. M. M. WINN. The Reaction $^{9}Be(d, ^{3}He)^{8}Li$ 946

Letters to the Editor:

Mr. M. M. WINN. Short-Lived Alpha Emitters produced by ^{3}He and Heavy Ion Bombardments 949

Dr. VACHASPATI and Mr. S. M. SHAH. Consistency of Nuclear Radii from Electron Scattering and Coulomb Energy Data 950

Reviews of Books 952

Contents of Section B 956

Field Aberrations in Wide Aperture Optical Systems

By W. S. S. BLASCHKE
Cooke, Troughton and Simms Ltd., York

Communicated by E. W. Taylor; MS. received 5th April 1954

Abstract. It is shown how the considerations which lead to the sine condition may be simply extended to take into account the aberrations which depend on the square of the field but include all powers of the aperture. In this way the relations known in primary aberration theory between pupil and image aberrations are reinterpreted and extended. The formulae are used to evaluate the field aberrations of a wide angle high numerical aperture flat field objective and comparison is made with the result of accurate trigonometrical calculation.

§ 1. Introduction

An important treatment of the optical sine condition has been given by Hopkins (1946), and is quoted in its essentials in Hopkins (1950). It is the aim of the present paper to show how this theory can be extended to obtain the aberrations occurring at larger fields. We shall simplify the treatment by considering unaberrated objects, the modifications which take into account the presence of aberrations in the object being in all cases immediately obvious.

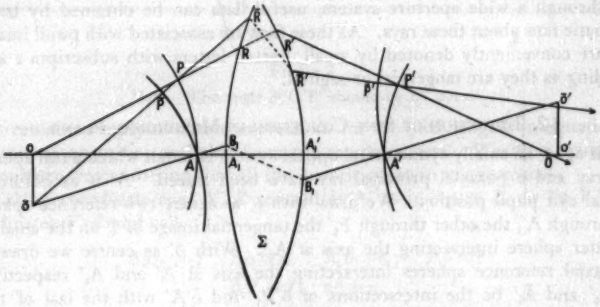


Figure 1.

Σ is an axially symmetrical optical system; A and A′ are the axial positions of entrance and exit pupil respectively; o and δ are axial and extra-axial object points; o′ and δ′ their respective gaussian images. Let reference spheres be drawn with each of these two pairs of points as centres to lie in the entrance and exit pupils respectively; $\delta AB_1 \ldots B_1'A'\delta'$ is a paraxial principal ray; $oPR \ldots R'P'O'$ is a full aperture axial (trigonometrically traced) ray and $\delta\bar{P}PR \ldots \bar{R}'\bar{P}'P'$ is taken to be a full aperture ray traced with the principal ray as axis, its aperture being determined by AP, where P is the intersection

of the axial ray with its reference sphere. P′, the intersection of R′O′ with $\overline{R}'\overline{P}'$ may be taken to lie on the axial reference sphere in image space when the fields considered are so small as to make the squares of the field angles negligible. The sine condition then asks what are the additional aberrations of the extra-axial ray taken relative to the principal ray over those of the axial ray at the same aperture. If oō is of the first order the paths of the two rays from P to P′ (outside and inside the optical system) can be considered neighbouring and by Fermat's theorem equal. The difference in aberration between the two rays, i.e. the change in optical path difference (O.P.D. axial minus ray path) reduces simply to $\Delta(\overline{P}P) = \overline{P}'P' - \overline{P}P$ (Hopkins 1950). This is the difference of the lateral defocusing terms in wavefront aberration theory, i.e. the tilt of a reference sphere in the exit pupil whose centre is moved from o′ to ō′, minus the corresponding term in object space. In this way the sine condition of Abbé or better the Stäble–Lihotzky condition is obtained.

The above treatment breaks down when the object and image heights (or other field parameter) are no longer considered small enough for the second powers to be neglected. The optical paths between P and P′ along oPR ... R′P′O′ and $\bar{o}\overline{P}\overline{R} \ldots \overline{R}'\overline{P}'$ can then no longer be considered equal. If on the other hand the two rays are viewed as constituting a second order fan imaging the point P (the edge of the entrance pupil) with the object height limiting the aperture, one sees that the optical path change may be allowed for by tracing astigmatic fans from this point about the trigonometrically traced axial ray. The optical paths in image space are then measured from the astigmatic foci P_s' or P_t' according as the rays lie in the sagittal or tangential sections. In other words, having traced a series of full aperture axial rays (marginal, submarginal, zonal, etc.) through a wide aperture system, useful data can be obtained by tracing astigmatic fans about these rays. As these fans are associated with pupil imagery they are conveniently denoted by small, barred letters with subscripts t and s according as they are tangential or sagittal.

§ 2. Extension of Sine Condition in Meridional Plane

Σ is again an axially symmetrical optical system through which a full aperture axial ray and a paraxial principal ray have been traced. A′ is as before the paraxial exit pupil position. We draw with o′ as centre two reference spheres, one through A′, the other through P_t' the tangential image of P on the axial ray, the latter sphere intersecting the axis at A_t'. With ō′ as centre we draw the extra-axial reference spheres intersecting the axis at A′ and A_t' respectively. Let $\overline{p}_t'$ and $\overline{A}_t'$ be the intersections of $\bar{o}'P_t'$ and $\bar{o}'A'$ with the last of these reference spheres. We have for the axial optical path difference

$$\Delta(W) = W' - W = [AA_1 \ldots A_1'A_t'A'] - [PR \ldots R'P_t'P']$$

and for the extra-axial optical path difference

$$\Delta(V) = V' - V = [AB_1 \ldots B_1'\overline{A}_t'A'] - [\overline{P}P\overline{R} \ldots \overline{R}'\overline{P}_t'\overline{P}']$$

in which square brackets denote optical paths.

It will be shown in Appendix I, that to the order of accuracy consistent with the aberration terms to be obtained, we may consider the reference spheres to be at A_t' instead of A′, although the constancy of aberration with progressing

wave fronts is strictly valid only when the reference spheres are centred at one of the invariant foci, as defined by Hopkins (1952). (In figure 2, O′ would be such a focus.) For the new positions of the reference spheres we have

$$\Delta(W)=[\mathrm{AA_1}\ldots\mathrm{A_1'A_t'}]-[\mathrm{PR}\ldots\mathrm{R'P_t'}]$$

$$\Delta(V)=[\mathrm{AB_1}\ldots\mathrm{B_1'\bar{A}_t'}]-[\mathrm{\bar{P}P\bar{R}}\ldots\mathrm{\bar{R}'\bar{P}_t'}].$$

From these

$$\Delta(V)-\Delta(W)=[\mathrm{A_t'A'}]-[\mathrm{\bar{A}_t'A'}]+[\mathrm{\bar{P}_t'P_t'}]-[\mathrm{\bar{P}P}] \qquad \ldots\ldots(1)$$

when equal optical paths from P to $\mathrm{P_t'}$ and A to A′ are cancelled. In equation (1) the latter two terms are again the sine condition terms (Hopkins 1950), modified slightly by the more accurate conditions of the present treatment. The first two are the astigmatic, cum spherical aberration on principal ray, terms, depending on the square of the field but including all the powers of aperture.

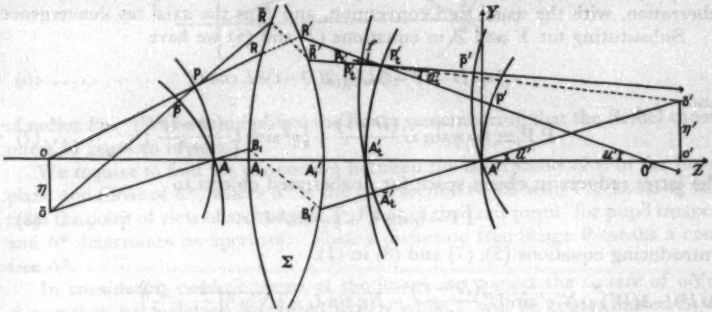


Figure 2.

Note. The angle A′O′P′ should be U' not u'

It remains simply to obtain expressions for the path lengths in equation (1) in terms of quantities available from the ray trace data. We drop primes for convenience, remembering that all the quantities are in image space.

Taking coordinate axes Y, Z centred at A′ as in figure 2, we let $\mathrm{P_t'}$ have coordinates Y, Z and call

$$\mathrm{A_to}\equiv r_t;\quad \mathrm{Ao}\equiv r;\quad \mathrm{\bar{A}_t\bar{o}}\equiv \bar{r}_t;\quad \mathrm{A\bar{o}}\equiv \bar{r}$$

and $$\mathrm{\bar{p}_tP_t}\equiv f;\quad \mathrm{o\bar{o}}\equiv \eta.$$

Then
$$r_t^2=Y^2+(Z-r)^2 \qquad \ldots\ldots(2)$$

$$\bar{r}_t^2=r_t^2+\eta^2. \qquad \ldots\ldots(3)$$

We have
$$(\mathrm{P_t\bar{o}})^2=(\bar{r}_t-f)^2=(Y-\eta)^2+(Z-r)^2$$

which by equations (2) and (3) reduces to

$$f^2-2f(r_t^2+\eta^2)^{1/2}+2\eta Y=0$$

from which, taking the smaller root of f and neglecting terms containing powers of η greater than the second we have

$$f=\eta Y/r_t+\tfrac{1}{2}\eta^2Y^2/r_t^3. \qquad \ldots\ldots(4)$$

As $\overline{P}_t P_t$ differs from f only by a quantity of $O((\lambda/ru)^2 f)$ where λ denotes an aberrational magnitude in wavelengths of light, $\Delta(f)$ expresses the comatic terms of the aberration change. For the first two terms on the right-hand side of equation (1) we have

$$\begin{aligned} A_t A - \bar{A}_t A &= r_t - r - (\bar{r}_t - \bar{r}) \\ &= -\tfrac{1}{2}\eta^2\left(\frac{1}{r_t} - \frac{1}{r}\right) \qquad \ldots\ldots(5) \end{aligned}$$

in which we have again neglected terms in powers of η greater than the second. To change equations (4) and (5) into formulae suitable for computation we have

$$Y = (D - \bar{l})\sin U; \quad Z = r - \delta L - (D - \bar{l})\cos U$$

where $\bar{l}$ and D are the intersection lengths along the axial ray from the last surface of the system to P_t and O respectively; $\delta L = oO$ is the longitudinal spherical aberration, with the usual sign convention, and U is the axial ray convergence.

Substituting for Y and Z in equations (2) and (4) we have

$$r_t^2 = (D - \bar{l})^2 + \delta L^2 + 2(D - \bar{l})\delta L \cos U \qquad \ldots\ldots(6)$$

and

$$\overline{P}\,P_t \simeq f = \eta \sin U \frac{(D - \bar{l})}{r_t} + \tfrac{1}{2}\eta^2 \sin^2 U \frac{(D - \bar{l})^2}{r_t^3}; \qquad \ldots\ldots(7)$$

the latter reduces in object space for unaberrated objects to

$$\overline{P}P = \eta \sin U + \tfrac{1}{2}\eta^2 \sin^2 U / r. \qquad \ldots\ldots(8)$$

Introducing equations (5), (7) and (8) in (1),

$$\begin{aligned} \Delta(V) - \Delta(W) = N'\eta' \sin U' \frac{(D' - \bar{l}')}{r_t'} - N\eta \sin U - \tfrac{1}{2}N'\eta'^2\left(\frac{1}{r_t'} - \frac{1}{r'}\right) \\ + \tfrac{1}{2}N'\eta'^2 \sin^2 U' \frac{(D' - \bar{l}')^2}{r_t'^3} - \tfrac{1}{2}N\eta^2 \frac{\sin^2 U}{r} \qquad \ldots\ldots(9) \end{aligned}$$

in which r_t is given from equation (6).

The extension of equation (9) to aberrated objects follows immediately, and the formulae are also easily modified for the case of telescopic systems.† Moreover, the result remains significant in the case of either infinite exit pupil, or finite gaussian pupil, but infinite tangential pupil image.

Equation (9) gives an extended form of relations between pupil and image aberrations familiar in Seidel theory. Thus from Wynne (1952) we have the equations

$$\Sigma\overline{S_V} - \Sigma S_{II} = H\Delta(u^2)$$

and

$$\Sigma\overline{S_{III}} - \Sigma S_{III} = H\Delta(u\bar{u})$$

where the notation is that of Hopkins (1950); $\frac{1}{2}S_V$, $\frac{1}{2}S_{II}$, $\frac{1}{2}S_{III}$ being the Seidel distortion, coma and astigmatism expressed as wavefront aberrations respectively.

It is known that there is a direct connection between the aberrations of an image point and those of the point on the imaging reference sphere through

† The astigmatism term $\frac{1}{2}N'\eta'^2\left(\frac{1}{r_t'} - \frac{1}{r'}\right)$ takes on the form $\frac{1}{2}N'\delta\bar{t}'\bar{u}'\bar{z}_t'$.

which the ray passes, when both are expressed as wavefront aberrations. Thus coma in the one expresses distortion in the other while both points will have the same astigmatic properties. This allows us to interpret the right-hand side terms of the above Seidel equations. By analogy we may then interpret the corresponding terms which appear in equation (9).

Dropping primes for convenience we again take the point P to be on the reference sphere centred at o and lying in the exit pupil. P itself will suffer from certain aberrations which are referred to a reference sphere Γ centred at P

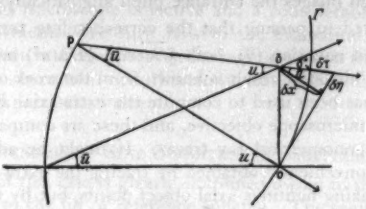


Figure 3.

of radius Po. (We have simplified the figure remembering that the Seidel theory refers to gaussian images.)

We require to find the connection between the aberrations of δ̄ in the image plane and those of δ̄*, where δ̄* is the intersection of Pδ̄ with Γ. Looking at it from the point of view of an image at P, Γ lies in the 'exit pupil' for pupil imagery and δ̄* determines its aperture. Then a distortion free image P means a coma free δ̄*.

In considering comatic terms of the image we neglect the square of oδ̄(η). ū may then be neglected compared with u while Γ will be approximated by its tangent plane. The connection between the aberrations of δ̄ and δ̄* require taking into account only the image shift δx considered parallel to Po as in figure 3. Expressed as a wavefront aberration this is

$$\tfrac{1}{2}Nu^2\,\delta x = \tfrac{1}{2}N\eta u^3 = \tfrac{1}{2}Hu^2$$

where $H = N\eta u$ is the Lagrange invariant. The object to image change in aberration is thus $\frac{1}{2}H\Delta(u^2)$ as required by the first of the above Seidel relations.

Consider next the astigmatism terms. An astigmatism-free image P means an astigmatism-free point δ̄*. To connect the astigmatism of δ̄ and δ̄* we require to take into account the additional displacements δη and δτ of figure 3, since ū may now no longer be neglected. Expressed as a wavefront aberration this gives

$$Nu\,\delta\eta - \tfrac{1}{2}Nu^2\,\delta\tau = Nu\eta u\bar{u} - \tfrac{1}{2}Nu^2\frac{\eta^2}{2r} = \tfrac{3}{4}Hu\bar{u}$$

to the required order of approximation. The object to image change in aberration is thus $\frac{3}{4}H\Delta(u\bar{u})$. This gives the appropriate correction term when it is remembered that the Seidel tangential astigmatism is given by $\frac{3}{4}S_{\mathrm{III}} + \frac{1}{4}S_{\mathrm{IV}}$ in which $S_{\mathrm{IV}} = \overline{S_{\mathrm{IV}}}$ is the Petzval curvature.

Since the astigmatism relation is simply reversed in sign when pupil and image are exchanged, the change in aberration due to the departure of Γ from

its tangential plane, i.e. $\frac{1}{2}H\Delta(u\bar{u})$ may equally well be considered in the exit pupil. Thus in equation (9) we convert $-\frac{1}{2}N\eta^2(r_t^{-1}-r^{-1})$ which expresses the reference sphere astigmatism into pupil astigmatism by the addition of a term of order $\frac{1}{2}H\Delta(u\bar{u})$. This adds to the last two terms of equation (9) which are the extended form of $\frac{1}{2}H\Delta(u\bar{u})$ to give the required correction term.

In many optical systems the correction terms in the astigmatism relation are small and may be neglected. It can then be said that an object is imaged free of tangential astigmatism (to the order η^2 but at all apertures) when a second order tangential fan images the entrance pupil stigmatically, i.e. when $r_t' = r'$.

It may be noted in passing that the corresponding term to $\frac{1}{2}\overline{S_V}$ does not appear explicitly in equation (9), both $\frac{1}{2}\overline{S_V}$ and $\frac{1}{2}H\Delta(u^2)$ being included in the sine condition. This latter result is known from the work of T. Smith.

Equation (9) has been used to compute the extra-axial aberrations of a high aperture flat field microscope objective, and these are compared with the results obtained from trigonometrical ray traces. It might be added that the latter results are most conveniently obtained by tracing the extra-axial rays, not as is usually done by taking fictitious axial object points, but by tracing rays directly relative to an already traced principal ray. The ray trace formulae used in this are an extension of an unpublished scheme developed by Dr. H. H. Hopkins for the axial case. The chief merit of the latter scheme and its extension lies in the fact that optical path difference is obtained without further calculation once the ray has been traced and checked.

The results are given in figure 4(*a*) together with a section of the objective.

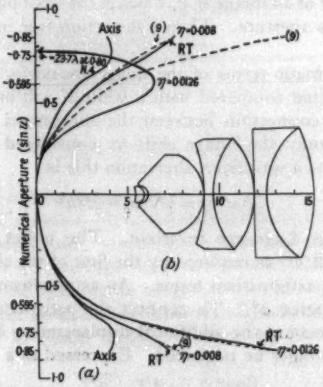


Figure 4. (*a*) Extra-axial wavefronts in meridional section of flat field objective. Paraxial (axial) wavefront coincident with *y*-axis. Object semi-fields $\eta = 0{\cdot}008$ and $0{\cdot}0126$ in. Full line represents ray trace values. Broken lines represent values computed from expression (9). (*b*) Section of 8 mm 0·85 numerical aperture flat field achromatic microscope objective. The abscissa denotes $\Delta(V)$ in units of λ.

At full field they show very close agreement on one side of the axis, but much less satisfactory agreement on the other. The reason for this is to be found in the presence of substantial amounts of higher order tangential coma or terms of comatic type. At the front of the objective, we have on one side of the axis the

addition of large angles of incidence for both axial and principal rays; considerable aberrations are bound to result in rays which are not far from being totally internally reflected. Nevertheless, a useful measure of field aberrations in an objective of large aperture and field is obtained. From the sine condition in its usual form, the objective was found to be coma-free to within one wavelength over the entire aperture at full field. Much of the additional aberration found by ray trace is thus accounted for by the calculations of terms in equation (9). The calculations have been repeated for a later and much more balanced correction of a similar flat field microscope objective and a considerable improvement in the agreement is evident. We turn now to the evaluation of the basic field curvature of the objective as it appears in the sagittal section free from odd power terms of wavefront aberration. As this normally involves the tracing of skew rays, the result is of particular value.

§ 3. Extension of Sine Condition in Sagittal Plane

The treatment of this case is greatly simplified owing to the vanishing of the comatic terms due to the tilt of reference spheres. The sagittal sections of the spheres are in fact coincident and only the astigmatism of pupil imagery remains.

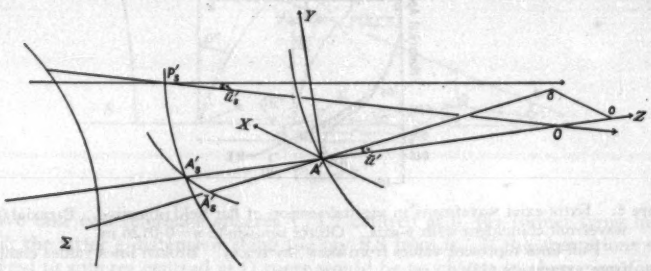


Figure 5.

We take as in the previous paragraph the axial reference sphere not in the exit pupil but intersecting the axial ray in P_s', the sagittal image of P. Let it intersect the axis at A_s' and the paraxial principal ray at $A'\bar{o}'$ at $\bar{A}_s'$. The change in O.P.D. extra-axial minus axial is

$$\Delta(V)-\Delta(W)=[A_s'A']-[\bar{A}_s'A'] \qquad \ldots\ldots(10)$$

and by an exactly analogous treatment to the above this becomes

$$\Delta(V)-\Delta(W)=-\tfrac{1}{2}\eta'^2\left(\frac{1}{r_s'}-\frac{1}{r'}\right) \qquad \ldots\ldots(11)$$

where

$$r_s'^2=(D'-\bar{s}')^2+\delta L'^2+2(D'-\bar{s}')\,\delta L'\cos U'. \qquad \ldots\ldots(11a)$$

The conversion of equation (11) to pupil aberration involves the addition of a term of order $\frac{1}{4}H\Delta(u\bar{u})$ all the other correction terms disappearing in the sagittal section. This corresponds to the relation involving the Seidel sagittal astigmatism which is $\frac{1}{4}(S_{III}+S_{IV})$.

The condition for freedom from sagittal astigmatism is then accurately $r_s' = r'$ and from equations (9) and (11) we have the vanishing of astigmatism when $r_s' = r_t'$ and $\frac{1}{2}\Delta\{N\eta^2 \sin^2 U(D-\bar{t})^2/r_t^3\}$ is negligible.

Equation (11) has been used to compute the extra-axial wavefront in the sagittal section of the same flat field microscope objective and the result compared with that obtained from skew traces (figure 6). The latter were used to compute the appropriate formulae of Hopkins (1952). It might be added that the shortest join of extra-axial ray to principal ray in image space at full aperture and field was found to be 0·010 in. This degree of skewness clearly illustrates the limitations of the sine condition when appreciable fields are considered.

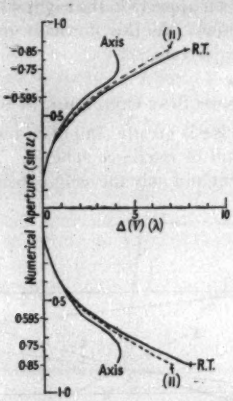


Figure 6. Extra-axial wavefronts in sagittal section of flat field objective. Paraxial (axial) wavefront coincident with y-axis. Object semi-field $\eta = 0{\cdot}0126$ in.
Full lines represent values from skew ray traces. Broken lines values computed from expression (11).

Finally it will be evident that the above approach can also be used to yield a simple form for the evaluation of distortion terms depending on large apertures. This would mean the computation of the pupil coma using data from the trigonometrically traced axial ray in an expression such as the one given in the last term of equation (106), Ch. VI (Hopkins 1950). This result may prove of value in photographic lens design.

Acknowledgments

Acknowledgment is made in the first place to Dr. H. H. Hopkins for a number of suggestions, in particular the method of deriving the transverse focal shift formulae given in Appendix II. Thanks are also due to Dr. G. C. Alway of the National Physical Laboratory for the tracing of skew rays. Dr. Alway asks it to be mentioned that these ray traces were done on the pilot A.C.E. electronic computor which once suitably programmed was able to trace a skew ray through ten surfaces of the system in about three seconds. Acknowledgment is finally made to Miss McAndrew for assistance in computations and to Messrs. Cooke, Troughton and Simms Ltd. for permission to publish.

APPENDIX I

The order of change in aberration at the gaussian focus as the wavefront progresses may be obtained for the general case. Let RS be a general ray through the optical system which does not necessarily intersect the axis Ao. Let sS be their shortest join. Then O, the midpoint of sS is an invariant focus, i.e. a point at which the aberrations expressed as a departure of the wavefront from the reference sphere are constant for all positions of the wavefront (Hopkins 1952). Moreover, any point on the skew bisector OQ will be an invariant focus and equivalent to O while on the other hand the particular aberrations of the gaussian image point must be related to the wavefront in the exit pupil. If we draw the perpendicular to the axis at o (the gaussian focus) to intersect the line of invariant foci at Q and let oQ be of length $\delta\tau$ then the transverse focal shift $\delta\tau$ will be the general equivalent of the longitudinal focal shift from ray to paraxial focus for intersecting rays.

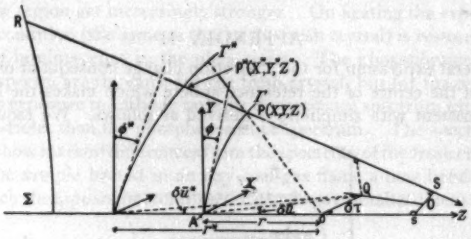


Figure 7.

We take two reference spheres Γ, Γ* centred at o, the former lying in the pupil, the latter a distance σ along the ray RS from it. If the aberrations were referred to spheres centred at Q there would be no change in aberration for the two positions of the wavefronts. It follows that the change in aberration of the gaussian focus as Γ moves to Γ* is equal to the difference in the defocusing terms due to the shift $\delta\tau$ taken for each wavefront in turn.

If now coordinate axes are so taken that oQ is along the X direction we have (see Appendix II)

$$\delta W_\tau^* - \delta W_\tau = \frac{X^*\delta\tau}{r^*} - \frac{X\delta\tau}{r} \quad \text{(neglecting terms dependent on } \delta\tau^2\text{)}$$

$$= X^*\delta\bar{u}^* - X\delta\bar{u}. \qquad \ldots\ldots \text{(A1)}$$

Assuming a distortion-free image of the reference sphere in the exit pupil by the ray RS the radius oP of the reference sphere Γ will have a convergence angle u and will lie along the paraxial ray. RS will depart from this by an angular aberration δu. Hence X will differ from Xr^*/r by a quantity depending on $\sigma\,\delta u$, viz. the projection of $\sigma\,\delta u$ along the X-direction. Thus

$$\delta W_\tau^* - \delta W_\tau = \mathrm{O}(\sigma\,\delta u\,\delta\bar{u}). \qquad \ldots\ldots \text{(A2)}$$

If ρ is the radial aperture of the wavefront in the exit pupil, the order of the angular aberration δu is given by the relation $\partial(\delta W)/\partial\rho = \delta u$ in which δW is a wavefront deformation of the order λ, where λ denotes the wavelength of light.

The order of $\delta\bar{u}$ is obtained from the fact that the change in aberration with transverse focal shift $\delta\tau$ is equivalent to the usual correction term (of order λ) introduced when the optical path difference computed for the ray focus is to be referred to the gaussian focus. Thus $Nu\,\delta\tau = O(\lambda)$ or $\delta\bar{u} = O(\lambda/\rho)$.

The quantity σ occurs in figures 2 and 4 of the previous paragraphs as $\delta\bar{l}$ and δs. It will in general be small. Should it however become very large in the case of either infinite gaussian or tangential (sagittal) exit pupil position, then the expression for the order of magnitude given in (A2) will nevertheless remain significant as in these cases $\delta\bar{u}$ or $\delta\bar{u}^*$ become vanishingly small.

Excepting these two cases we may thus write

$$\delta W_\tau^* - \delta W_\tau = O\left(\frac{\delta\bar{l}}{\delta\bar{s}}\left(\frac{\lambda}{\rho}\right)^2\right) \qquad \ldots\ldots(A3)$$

which justifies the considerations of §§ 2 and 3.

APPENDIX II

The general expression for the aberration change consequent on a transverse focal shift of the centre of the reference sphere which makes the least approximations consistent with simplicity is derived as follows. We require the value

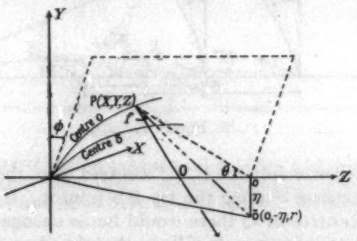


Figure 8. The origin of the coordinates is at the axial pupil point A.

of the interval f between the two reference spheres centred at o and ō respectively and are justified as above in taking it along the radius of the reference sphere centred at ō of radius ōA and not along the ray. Then

$$(P\bar{o})^2 = \{(r^2+\eta^2)^{1/2} + f\}^2$$

but

$$(P\bar{o})^2 = X^2 + (Y+\eta)^2 + (Z-r)^2$$

from which we obtain by the use of the relation $X^2 + Y^2 + (Z-r)^2 = r^2$ a quadratic in f: $f^2 + 2(r^2+\eta^2)^{1/2} f - 2Y\eta = 0$, the smaller root of which, to the order η^2, is $f = Y\eta/r + \frac{1}{2}Y^2\eta^2/r^3$. Then since a negative η represents a tilt of the reference sphere in the positive sense (i.e. in the direction of the light ray),

$$\delta W_\eta = -Nf = -N\eta \sin\theta\cos\phi - \tfrac{1}{2}N\eta^2 \sin^2\theta\cos^2\phi/r. \qquad \ldots\ldots(A4)$$

References

HOPKINS, H. H., 1946, *Proc. Phys. Soc.*, **58**, 92; 1950, *The Wave Theory of Aberrations* (Oxford: University Press); 1952, *Proc. Phys. Soc.* B, **65**, 934.

WYNNE, C. G., 1952, *Proc. Phys. Soc.* B, **65**, 429.

Cathodo-Luminescence of MgO

By B. D. SAKSENA† and L. M. PANT‡
† National Physical Laboratory of India, New Delhi, India
‡ Department of Physics, University of Allahabad, Allahabad

MS. received 8th June 1954

Abstract. The cathodo-luminescence spectra of different samples of MgO—fresh, previously exposed to cathode rays, heated in a Bunsen and oxy-coal-gas flame etc.—show different characteristics. For the fresh crystal we get a strong band with peak at 4550–4400 Å which weakens in intensity as the crystal receives increasing previous exposure to cathode rays. At the same time three bands in the red–yellow region get increasingly stronger. On heating the exposed crystal, the previous condition (the same as that of the fresh crystal) is restored except for an increase of luminescence in the ultra-violet. The phosphorescence and the fluorescence spectra are the same for the fresh crystal, but for the crystal heated after previous exposure to cathode rays the fluorescence spectrum extends farther into the ultra-violet than the phosphorescence spectrum. The spectra of heated samples also show marked differences from the spectrum of the fresh crystal, and in the case of the sample heated in an oxy-coal-gas flame a new band at 2450 Å is obtained which disappears on continued cathode-ray bombardment.

§1. Introduction

The existence of negative-ion vacancies and the formation of colour centres in alkali halides has been established from the work of Hilsch and Pohl (1937) and other workers. Their presence has also been suspected in MgO on account of the decay observed in secondary emission during cathode bombardment (Wright 1950, see also Hibben 1937). Recently Weber (1951) has produced spectroscopic evidence for their existence by obtaining absorption bands at 5·7 and 4·34 ev from MgO with an excess of oxygen and at 4·9, 3·6 and 2·35 ev from MgO with an excess of metal which can be bleached by heating the crystals *in vacuo* at 1500°K. In the present paper we support the work of Weber by a study of the cathode luminescence spectra of MgO. We have studied the emission from the crystal and from the light MgO powder, the effect of previous heat treatment and previous exposure to cathode rays on the cathodo-luminescence, and the phosphorescence spectrum.

So far as we are aware no detailed study of the cathodo-luminescence spectra of MgO has previously been made. Northrop (see Nichols, Howes and Wilber 1928) obtained a broad band between 5230 and 3000 Å with a maximum at 4050 Å for MgO I and a band between 5280 and 3090 Å with a flat maximum between 3970 and 3510 Å for MgO II, but he does not state the meaning of the figures I and II. Deutschbein (1932) and Ewles (1938) have studied cathode luminescence spectra with different activators while Bube and Stripp (1952) find a faint blue luminescence in MgO without added impurity.

§ 2. Experimental

The crystalline specimens used were clear cleavage pieces from a big crystal of MgO. Another form of MgO powder called light MgO, prepared by Messrs Boots Pure Drug Co., Nottingham, England, was also studied. X-ray photographs of both powders gave the same value of 2·11 Å for the shortest Mg–O distance though the rings of the light powder were diffuse and could be obtained only after long exposures. We have examined the spectrum of the crystal in an arc with carbon and copper electrodes and feel that traces of calcium might be present in the crystal as an impurity.

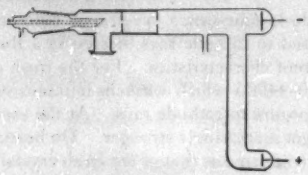


Figure 1.

The cathode luminescence tube is shown in figure 1. It was exhausted for several hours by two Warren mercury diffusion pumps backed by a Cenco pump, and before the experiment was started the tube was degassed several times and the vacuum tested to see that it had reached the x-ray stage (10^{-5}–10^{-4} mm). The tube was run at 40 000 volts obtained with the help of a large induction coil. The crystal was fitted in a small depression in the anticathode and was exposed directly to the beam of cathode rays since the anode lay in the side tube. The possibility of the excitation of the crystal by ions does not appear to be large because at such low gas pressures and with electrons of high energies the number of stable negative ions formed will be quite small and the positive ions will move away from the crystal. It may be argued that the resonance Mg I line at 2852 Å recorded by us in some spectrograms of plate II is due to the release of Mg atoms from the surface of the crystal by ion bombardment, but excitation for a long period by cathode rays of large energy will also produce free Mg atoms on the surface of the crystal, and this would explain the absence of the 2852 line in the spectrum of the fresh crystal and its large intensity in the spectrum of the crystal having large previous exposure to cathode rays (spectrograms v and vi, plate II). Apart from these considerations the bands with which we are primarily concerned in this paper are broad and structureless and their intensity changes considerably with the previous treatment of the crystal. It is therefore justifiable to assume that they represent the luminescence spectrum of the crystal.

A small quartz spectrograph (dispersion 125 Å mm^{-1} near 4300 Å and 285 Å mm^{-1} near 5750 Å) and a constant deviation glass spectrograph (dispersion 23 Å mm^{-1} and 68 Å mm^{-1} in the same regions) were used for photographing the spectra. The plates used were Kodak super panchro-press plates whose spectral sensitivity in the region 5700–6500 Å was tested by taking the spectra of the radiations emitted by a globar lamp and an electric bulb. In both cases the plates showed a continuous variation of intensity without any banded structure except for a narrow region of increased sensitivity close to 5782 Å. The plates were sensitized by the method

described by Dundon, Schoen and Briggs (1926) for use in the region 6500–8000 Å. The spectra were microphotometered on a Hilger H451 non-recording instrument.

For photographing the phosphorescence spectrum it is necessary to allow no light to enter the spectrograph when the crystal is under cathode bombardment. It was therefore arranged that before the knob of the commutator of the induction coil reached the position where the cathode rays were turned on, a strong electromagnet was energized and attracted a soft iron piece so as to cover the slit of the spectrograph completely. For this purpose the ebonite knob was provided with a metal screw which, on the turning of the knob, made contact with a metal plate below and brought into operation a local battery connected to the electromagnet. Since the phosphorescence was very weak, it was necessary to excite the electromagnet repeatedly at short intervals and to give an exposure of 25 hours with a fairly wide slit on the small quartz spectrograph.

§3. Results

(*a*) *Effect of Previous Exposure to Cathode Rays on Cathode Luminescence*

In the spectrum of the fresh crystal (spectrogram *a*, plate I) the band 5050–4050 Å is the most intense and has a maximum at 4550–4400 Å, while the band 5850–5650 Å is also weakly present. The intensity in the region 5850–5650 Å may also be due to the increased sensitivity of the plate in this region and the presence of two mercury lines. With a quartz spectrograph the former band was found to extend to 3350 Å (spectrogram vi, plate II). It may be noted from spectrograms *a*, *b* and *c* of plate I and the microphotometer curves *a* and *b* of figure 2 that with greater and greater previous exposure to cathode rays the band

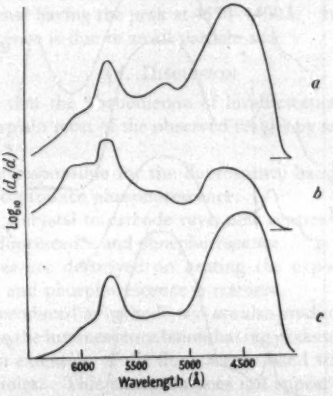


Figure 2. Density wavelength curves of the cathodo-luminescence spectra. *a*, fresh MgO crystal; *b*, crystal exposed previously to cathode rays for 30 hours; *c*, crystal heated in bunsen flame after 30 hours previous to cathode rays and cooled.

5050–4050 Å appears to lose in intensity while the bands 6500–6380, 6150–5950, and 5850–5650 Å increase in intensity. Further, from the spectrograms v and vi of

plate II it is seen that the spectrum of the crystal previously exposed to cathode rays extends farther into the ultra-violet than that of the fresh crystal.

The crystal that has been exposed to cathode rays turns brownish but the colour disappears on heating it strongly. The difference in the spectra of these two types is shown by spectrograms *c'* and *d* of plate I. It may be noted that the two bands at 6500–6380 Å and 6150–5950 Å in the spectrogram *c'* are almost absent in *d*, while the band 5050–4050 Å of spectrogram *c'* becomes the strongest band in *d*. Also the spectrogram *d* appears the same as spectrogram *a* of the fresh crystal. These observations are borne out by microphotometer curves *a* and *c* of figure 2. If this crystal (i.e. the one heated after previous exposure to cathode rays) is further exposed to cathode rays we find changes similar to those observed in a fresh crystal on exposure to cathode rays (spectrograms iv and v, plate II).

A band at 6775–7790 Å is present very faintly in the spectrum of a fresh crystal heated in a Bunsen flame (spectrogram viii, plate II), but quite prominently in the spectrum of the crystal previously exposed to cathode rays (spectrograms ix and x, plate II).

(*b*) *Effect of Heat Treatment on Cathode Luminescence*

In the spectrum of the fresh crystal heated in a Bunsen flame (spectrogram iii, plate II) we notice that the band with the peak at 4550–4400 Å extends to 3200 Å while the bands in the red–yellow region are quite prominent though weaker than the former band. The difference between this spectrum and that of the fresh crystal (spectrogram vi, plate II, and microphotometer curves *f* and *g*, figure 3) is

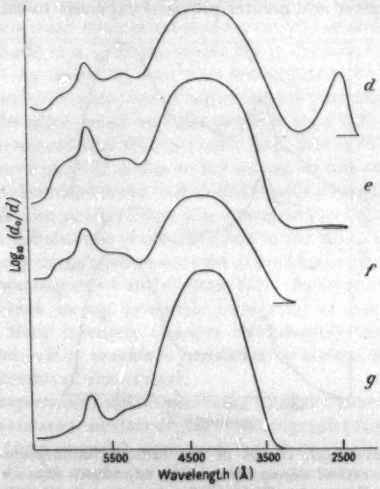


Figure 3. Density wavelength curves of the cathodo-luminescence spectra. *d*, MgO crystal heated in oxy-coal-gas flame for 3 min and cooled; *e*, same crystal as *d* subsequently exposed to cathode rays for 15 hours; *f*, fresh crystal heated in Bunsen flame for 5 hours and cooled; and *g*, fresh crystal.

quite marked because in the latter case the bands in the red–yellow region are very faint compared with the strong band with the peak at 4550–4400 Å, which extends to 3350 Å.

The spectrum of the fresh crystal heated in an oxy-coal-gas flame (spectrogram i, plate II, and microphotometer curve *d*, figure 3) shows a new band between 2850 and 2250 Å with a maximum at 2450 Å besides those observed in the spectrum of the crystal heated in a Bunsen flame. The band, however, disappears when the crystal has been exposed to cathode rays for several hours (spectrogram ii, plate II, and microphotometer curve *e*, figure 3).

The bands in the red–yellow region have an appreciable intensity in the spectra of heated crystals, which suggests that the effect of intense heat treatment is similar to that of previous exposure to cathode rays.

(*c*) *Phosphorescence Spectra*

The phosphorescence is largely associated with the band having peak at 4550–4400 Å. The fluorescence and phosphorescence spectra are the same for the fresh crystal (spectrograms *e* and *f*, plate I). But for the crystal heated after previous exposure to cathode rays one finds a larger extension towards the ultra-violet for the fluorescence spectrum, though the phosphorescence spectrum is the same as that of the fresh crystal (spectrograms *g* and *h*, plate I).

(*d*) *Emission Spectrum of the Crystal Heated in the Oxy-Coal-Gas Flame*

The spectrum extends from 6700–3400 Å with its maximum in the red. The glow was dazzling white.

(*e*) *Luminescence of Light Powder*

The luminescence is extremely faint though it lies in the same region as the band of the fresh crystal having the peak at 4550–4400 Å. It is possible that the faintness of luminescence is due to small particle size.

§ 4. Discussion

It is well known that the phenomenon of luminescence is due to impurity centres. One can explain most of the observed results by assuming four centres A, B, C and D.

(i) Centres A are responsible for the fluorescence band having the peak at 4550–4400 Å, and also produce phosphorescence.

(ii) On exposing a crystal to cathode rays, new centres B are created which quench the original fluorescence and phosphorescence.

(iii) These centres are destroyed on heating the exposed crystal, and the original fluorescence and phosphorescence is restored.

(iv) The centres produced by cathode rays are also produced by heating a fresh crystal and give rise to the luminescence bands having peaks at 2·16, 2·04, 1·92 and 1·73 ev and also to an extension of the fluorescence band with the peak at 4450–4400 Å in the ultra-violet. This extension does not appear in phosphorescence, and it is possible that it is due to another centre C.

(v) On heating a fresh crystal in an oxy-coal-gas flame a new centre D is produced which gives rise to the luminescence band having peak at 2450 Å (5·06 ev).

No definite conclusions can be drawn about the nature of these impurity centres from the present data, but it is possible that the B centres that are found in

MgO on cathode excitation may be metal-ion-excess impurity centres, while the D centres that are formed when the crystal is heated in an oxy-coal-gas flame may be negative-ion-excess impurity centres. This view is in agreement with the well-established ideas regarding the formation of F and V colour centres in alkali halides; and it is satisfactory to find that the band associated with the D centre lies to the short wavelength side of the bands associated with the B centre.

ACKNOWLEDGMENT

Our thanks are due to the Council of Scientific and Industrial Research for a research grant and a scholarship to one of us (L.M.P.).

REFERENCES

BUBE, R. H., and STRIPP, K. F., 1952, *J. Chem. Phys.*, **20**, 193.
DEUTSCHBEIN, O., 1932, *Ann. Phys., Lpz.*, **14**, 712, 729; 1932, *Z. Phys.*, **77**, 489.
DUNDEN, M. L., SCHOEN, A. L., and BRIGGS, R. M., 1926, *J. Opt. Soc. Amer.*, **12**, 397.
EWLES, J., 1938, *Proc. Roy. Soc.* A, **167**, 34.
HIBBEN, J. H., 1937, *Phys. Rev.*, **51**, 530.
HILSCH, R., and POHL, R. W., 1937, *Proc. Phys. Soc.*, **49**, 3.
NICHOLS, E. L., HOWES, H. L., and WILBER, D. T., 1928, *Cathode Luminescence and Luminescence of Incandescent Solids* (Carnegie Institute of Washington), p. 106.
WEBER, W., 1951, *Z. Phys.*, **130**, 392.
WRIGHT, D. A., 1950, *Semi-conductors*, Methuen monograph, p. 88.

Luminescence Characteristics of Some Scintillating Crystals

By J. R. COOK and K. A. MAHMOUD

Department of Physics Applied to Medicine, The Middlesex Hospital Medical School, London

Communicated by J. E. Roberts; MS. received 9th November 1953, and in amended form 25th June 1954

Abstract. The phosphorescence observed during the period 2 to 400 μsec following the absorption of gamma-ray photons has been studied in crystals of sodium iodide and potassium iodide (thallium activated) and calcium tungstate. The integrated light output is found to vary with temperature over the range -180°C to $+15$°C in a complex manner differing with each crystal. On the other hand, the decay time of the after pulses is to a first approximation temperature independent, with a decay constant in these particular crystals of around 15 μsec. It is concluded that a temperature dependent dissipative process occurs at some time during the initial decay, and is followed by the decay to a ground state with a time constant independent of temperature.

§ 1. Introduction

The existence of metastable states in the alkali halides has been shown by various experimenters (Bünger and Flechsig 1931, Randall and Wilkins 1945). Early experiments have used phosphoroscopes and photoelectric cells to measure the decay time and light intensity as a function of temperature. Later experimenters (Smaller, May and Freedman 1950, Bonanomi and Rossel 1951) have used photomultipliers and counting systems, and have in general confirmed the theoretical treatment of thermoluminescence which was first suggested by Randall and Wilkins (1945).

The use of thallium activated sodium iodide crystals for scintillation counting was initiated by Hofstadter (1948) and many other crystals are now used for this purpose. Some of them show a phosphorescent effect and several experiments have been done on the after pulses which occur in the commonly used crystals (e.g. Cross 1950). Strange (1946) has made detailed investigations of the emission processes in cathodoluminescence with low energy electron beams on various phosphors.

In the present work we have used the scintillation technique of crystal and photomultiplier to analyse the after pulses in greater detail, and have investigated their variation with temperature. Previous experiments using a similar technique differ in an important aspect from the present one, in that they recorded the pulses which appeared after an intense gamma irradiation, followed by warming. They were, therefore, concerned with thermoluminescence. In the present work we have measured the after pulses from primaries which occur coincidentally with the passage of gamma rays through the crystal, as a function of the crystal temperature. We are concerned therefore, with the immediate decay mechanism when energy is transferred from the secondary electron to the crystal.

§ 2. Experimental Method

2.1. *Decay Measurement*

The crystal was placed in a cylindrical cavity inside a cone-shaped piece of Perspex (see figure 1). The optical efficiency was increased by a paraffin oil filling and a silver coating on the cone. The crystal was then cooled by surrounding the end of the cone with liquid nitrogen, its temperature being measured by an adjacent thermocouple. When the liquid nitrogen was removed the temperature of the crystal increased at a rate of approximately 200°c in 15 minutes so that measurement of the scintillation pulses could be made over particular regions of temperature.

A source of ^{60}Co was used to produce the light flashes which were recorded by an EMI 5311 photomultiplier and a linear amplifier, the latter having a rise time of approximately 0·15 μsec and a resolving time of approximately 2 μsec. Photographic registration of the output pulses was used, with a high speed synchroscope, to obtain a permanent record. Figure 2 (Plate) gives two examples of the recorded trace. The separation of the emission into a

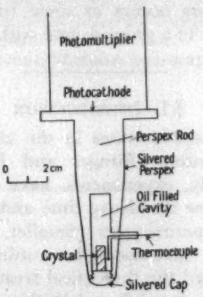


Figure 1. Crystal mount.

primary pulse and secondary pulses is obvious from figure 2 (*a*). It was necessary to use a small time constant at the amplifier input so that the secondary pulses could be clearly separated. If the integration time constant had been much smaller then individual photoelectrons in the pulses would have registered. When the resolution of the secondary pulses was incomplete an approximate correction was made in measuring amplitudes. In general, measurements were not made between 0 and 5 μsec after the primary pulse because of the superposition of secondaries and the difficulty in defining their amplitude. The apparent variation in the base line in figure 2 (*a*) is due to the ' back kick ' of the amplifier and pulse heights have been measured above this base line.

The circuit actuating the synchroscope was an integral discriminator. Only those primary pulses arising from electrons lying approximately between 1 and 1·33 MeV were recorded from each crystal, the maximum pulse height corresponding to the 1·33 MeV photoelectric absorption. Occasional large pulses were attributed to the coincidence of cascaded gamma rays. Because this

method excluded small pulses it was most likely that all the observed pulses occurred simultaneously with β-rays in the crystal and were not due to phosphorescent decays from metastable states such as were observed by Bonanomi and Rossel (1951).

The primary pulse height corresponding to the photoelectric peak was determined by photographing the envelope of the pulses on a faster sweep with reduced discriminator voltage. This can clearly be seen from figure 2 (*b*), which shows the change in peak primary height with temperature.

Particular care was taken to ensure that no spurious pulses occurred which could be attributed to the photomultiplier. This point was checked by irradiating the Perspex cone with the ^{60}Co source in the absence of a crystal. The Perspex produced primary scintillation pulses which had no observable secondaries over the temperature and time ranges used in the experiment. An anthracene crystal under the same conditions also showed primary pulses only. Some of the observations were also repeated with a second photomultiplier which gave results consistent with those described here. The difference between the light emission–temperature curves for each crystal could hardly be attributed to the photomultiplier.

2.2. *Fluorescent Spectra*

In order to facilitate comparison with other workers, the emission spectra of our crystals, irradiated with ultra-violet light, were recorded on a Hilger spectrograph. Peak wavelengths and bandwidths are shown in the table.

Crystal	Peak (Å)	Bandwidth to 50% peak int. (Å)
$CaWO_4$	4200	750
NaI (Tl)	3800	740
KI (Tl)	4250	1000

§ 3. Method of Analysis

Each photograph was projected on to a scale so that the height of each secondary and its position on the time scale relative to the primary could be measured. While the height scale was in arbitrary units (which were the same for the envelope heights), the time scale was fixed accurately by a pulse generator.

Let N_h secondary pulses occur with a height between h and $h+\Delta h$ in a time after the primary between t and $t+\Delta t$.

Then the number $\Sigma_h N_h h$ plotted as a function of t will represent an integrated curve of total phosphorescent intensity as a function of time, at a particular temperature T°c. In order to determine such a curve with reasonable accuracy it was necessary to measure several hundred pulses at each temperature.

Let N_p be the number of primary pulses at T°c which have at least one secondary. Then

$$I_T = \frac{\Sigma_h \Sigma_t N_h h}{N_p}$$

is proportional to the average light output per primary from those secondaries in the region 2 to 400 μsec at T°c.

Let N_p' be the total number of primary pulses at T°c (which includes

those pulses having no secondaries). Now N_p' includes a small number of pulses from the Perspex cone which are unaccompanied by secondaries. Because we had no arrangement for determining the film speed we could not determine this number with accuracy. Graphs of $I_T' = (\Sigma_h \Sigma_t N_h h)/N_p'$ differ only slightly in form from those of I_T/T, particularly at low temperatures (figures 5, 6, 7 and 8). At very low temperatures (-180°c) all of the crystals showed a marked decrease in the number of secondary pulses. The probable errors of some points are indicated.

§ 4. Experimental Results

4.1. *Decay of Secondaries*

In the case of $CaWO_4$, NaI (TlI) and KI (TlI) the graphs of $\Sigma_h N_h h$ against time were found to be approximately exponential and at each temperature the best value of the decay constant τ was found by the method of least mean squares. With the data having a higher statistical accuracy the value of the correlation coefficient discriminated in favour of an exponential rather than a linear decay law.

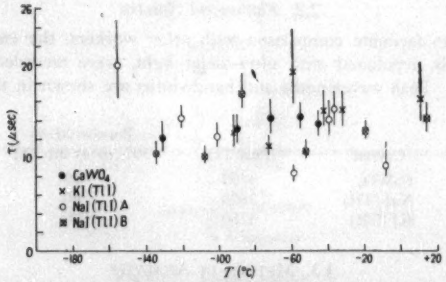


Figure 3. τ as a function of temperature.

Figure 3 shows τ as a function of T for the four crystals. The probable error of individual points is shown as determined from the pulse height distribution of secondaries in a given time interval. Points on the (τ, T) graph which have a large probable error result from the small number of secondary pulses per primary at that temperature. It is clear, however, from figure 3 that there is no marked or regular change of τ as a function of temperature. The average time constant of all the crystals is approximately 15 μsec.

A better check of an exponential decay is obtained if the experiment is done over several half-lives. Though the majority of measurements were made in the region 5 to 40 μsec, in certain cases we investigated the region 2 to 400 μsec by varying the time base. A decay curve for $CaWO_4$ at room temperature is shown in figure 4. In this case there is definite evidence of a long duration low intensity phosphorescence occurring following an exponential decay of the type described above. The experimental errors are large so we cannot say whether the long period decay follows an exponential or hyperbolic law. In

the case of NaI (Tl) and KI (Tl) such a long duration phosphorescence has not been observed.

In the case of $CaWO_4$ the value of τ was also measured with secondaries arising from 0·5 MeV primary Compton pulses as opposed to the approximately 1 MeV primary pulse size of the above data. The value of τ was approximately 15 μsec which suggests that τ does not vary with the initial electron energy.

4.2. *Average Light Output of Secondaries*

While the decay time τ remains constant the function I_T' shows a marked variation with temperature as can be seen from figures 5, 6, 7 and 8, which are discussed below.

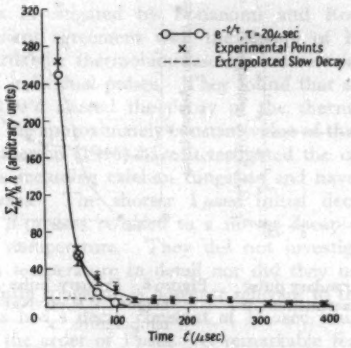


Figure 4. Decay of $CaWO_4$ phosphorescence from 0 to 400 μsec at 15°C.

4.3. *Primary Pulse Height*

This was determined by the method of envelope exposures and was found to be temperature dependent. Curves for the various crystals are shown on the same figures as for the secondary pulse output.

4.4. *Characteristics of Individual Crystals*

(*a*) *Calcium tungstate.*

This was in the form of a small cylinder of diameter 4 mm and length 0·5 cm (artificially grown by the Linde Air Co. of America) with polished surfaces. Figure 5 shows the marked increase of secondary emission I_T' between -60°c and -40°c. The percentage increase is obviously much greater than that in the primary height over the same temperature range, which is further evidence of the non-spurious nature of the pulses.

Between -60°c and -120°c the variation of I_T' against T is less rapid and at -180°c no secondaries were observed. The tendency of I_T' to reach a plateau may also be noted.

(b) Sodium iodide (Tl).

Two small crystals have been used containing a small but unknown percentage of thallium and of size approximately 1 cm × 0·5 cm × 0·5 cm. It is evident from figures 6 and 7 that the crystals (NaI (Tl) A and B) exhibit different characteristics.

Crystal A shows a single peak in the curve of I_T' against T at a temperature of $-70°$C, while the primary pulse increases in a marked way with temperature. On the other hand, crystal B shows two peaks on the (I_T', T) curve at $-75°$C and $-20°$C. In this case the primary pulse height exhibits a peak at $-90°$C.

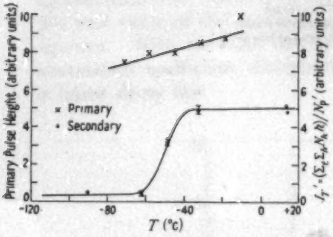


Figure 5. Primary and secondary pulse height of $CaWO_4$ as a function of temperature.

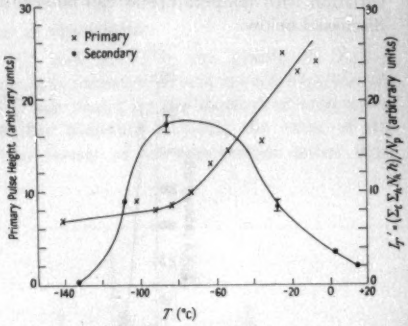


Figure 6. Primary pulse height and secondary pulse height ×2 of NaI (Tl)A as a function of temperature.

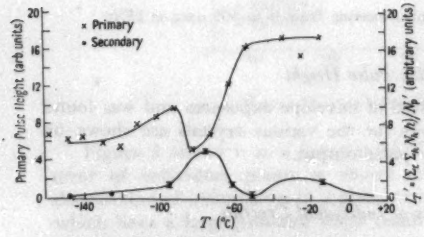


Figure 7. Primary and secondary pulse height of NaI (Tl)B as a function of temperature.

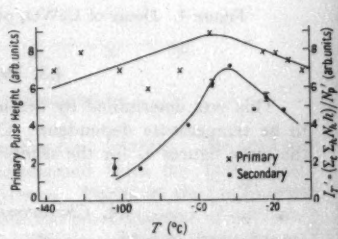


Figure 8. Primary pulse height and secondary pulse height ×10 of KI (Tl) as a function of temperature.

(c) Potassium iodide (Tl).

In this case the (I_T', T) curve shows a single peak at approximately $-50°$C (figure 8). Measurements of the envelope height were less accurate but indicate no marked change as a function of temperature. The average light output per primary pulse from the secondaries at room temperature is larger than in the case of the sodium iodide crystals.

§ 5. Conclusion

It is evident from the above work that the process which is responsible for the decay of order 15 μsec is different from that which constitutes the primary pulse. As the temperature decreases the percentage of light output due to the secondaries decreases until all the phosphorescence is included with the primary pulse. This means that at low temperatures (~ -180°c) the individual photons which make up the primary pulse must be grouped mainly in the region 0–2 μsec and the process which produces the later secondaries is quenched by some thermal mechanism. At higher temperatures the probability of a slow decay increases but the decay time remains approximately constant.

The alkali halides show definite peaks in the secondary emission–temperature curves and their decay time behaviour is strikingly different from the mechanism of metastable states investigated by Bonanomi and Rossel (1951). These workers found excellent agreement with the theory of Randall and Wilkins (1945) both as regards the thermoluminescence curves and the measurement of the decay time of individual pulses. They found that a temperature change from -50°c to -180°c altered the decay of the thermoluminescence by a factor 10^6, as against the approximately constant value of the present experiment.

Strange and Henderson (1946) have investigated the cathodo-luminescence of various phosphors including calcium tungstate and have found evidence for two separate processes. The shorter 1 μsec initial decay was termed the α-process while the β-process referred to a slower decay whose time constant was independent of temperature. They did not investigate the variation of these processes with temperature in detail nor did they use the alkali halides. Their result for calcium tungstate is however relevant to the present discussion in that the β-process had a decay constant of 15 μsec while the α-process had a decay constant of the order of 1 μsec. A remarkable feature of their results was the similar behaviour of many different phosphors whereby the β-process had a decay time of the same order as that of $CaWO_4$.

In the present experiments we have found evidence for a similar state of affairs in $CaWO_4$ and the thallium activated alkali halides. We are however dealing with the excitation by high-energy electrons and not with an electron beam of low energy.

The theoretical interpretation of these results is uncertain. It is necessary to postulate that the probability of excitation of states responsible for the β-process is decreased at low temperatures. The excited states responsible for the α-process are less affected by temperature. The results of Elmore and Hofstadter (1949) indicate some change of decay time with temperature in the interval 0–1 μsec.

The two processes may be associated with different energy levels in the same emissive centres. The peaks in the emission–temperature curves for the β-process may possibly represent the competing effect of exciton dissociation and non-radiative processes both of which would increase with temperature (see Mott and Gurney 1950).

Acknowledgment

We are grateful to Professor J. E. Roberts for helpful criticism of this work.

References

Bonanomi, J., and Rossel, J., 1951, *Helv. Phys. Acta*, **24**, 310.

Bünger, W., and Flechsig, W., 1931, *Z. Phys.*, **67**, 42.

Elmore, W. C., and Hofstadter, R., 1949, *Phys. Rev.*, **75**, 203.

Cross, W. G., 1950, *Phys. Rev.*, **78**, 186.

Hofstadter, R., 1948, *Phys. Rev.*, **74**, 100.

Mott, N. F., and Gurney, R. W., 1950, *Electronic Processes in Ionic Crystals* (Oxford: University Press).

Randall, J. T., and Wilkins, M. F. H., 1945, *Proc. Roy. Soc.* A, **184**, 366.

Smaller, B., May, J., and Freedman, M., 1950, *Phys. Rev.*, **79**, 940.

Strange, J. W., 1946, *Proc. Phys. Soc.*, **58**, 368.

Strange, J. W., and Henderson, S. T., 1946, *Proc. Phys. Soc.*, **58**, 392.

Out-of-Focus Diffraction Patterns for Microscope Objectives

By R. O. GANDY
University College, Leicester

MS. received 3rd December 1953

Abstract. Formulae are derived for the diffraction pattern at the image plane of an out-of-focus point source, and an out-of-focus line source. The lens considered is an idealization of a microscope objective; it is assumed to be a flat-field anastigmat, and to subtend a small angle at the image, but no restriction is placed on the angle subtended at the object. A discussion is given of the assumptions required both for deriving and for applying the formulae, and some indication is given of the effect of the differences between practice and theory.

THERE are two reasons why the classical results of Lommel (1885) and Struve (1886) cannot be applied to a microscope objective. Firstly, if the lens satisfies the sine condition, then the wave front which emerges on the image side when a point source is placed on the object side, will only be spherical if the source lies at the correct working distance from the lens: *an out-of-focus source produces a non-spherical wave front.* (So that the patterns produced at the correct image plane, as an object is moved out of focus, are not the same as those which will occur if the object is left at its correct position and the image plane is moved.) Secondly, the amplitude of the disturbance will not be uniform over the converging wave front.

In this paper we calculate the amplitude distribution over the image plane due to an out-of-focus point and an out-of-focus coherent line, taking into account the two facts just mentioned. We assume that the magnification of the lens is large, so that the angle it subtends at the image is small, although the numerical aperture of the lens may be large. We further assume that the lens is an ideal flat-field anastigmat—and hence, in particular, that it satisfies the sine condition. This assumption is formulated as follows: there are two particular planes, the *object* and the *image* planes, which are perpendicular to the axis of the system, and all the rays from any point P of the object plane are brought to a focus at the corresponding point P′ of the image plane; i.e. the phase delays along the various ray-paths from P to P′ are all equal.

Let the axis of the system be taken as the z axis, and let O and O′ be the points in which it meets the object and image planes. Let S and S′ be spheres with centres O and O′ which lie wholly on the object and image sides of the lens system, and let R be the radius of S. Let μ, μ' be the refractive indices, k $(=2\pi/\lambda)$, k' the propagation constants for the media on the object and image sides of the lens. Then the disturbance $V(\mathrm{Q})$ at a point Q on S due to an isotropic source at P in the object plane is given approximately by:

$$V(\mathrm{Q}) = B\exp(-ik\mathrm{PQ}),$$

where B is a constant. Let P have coordinates (x, y), and let Q have polar coordinates (R, θ, ϕ); then

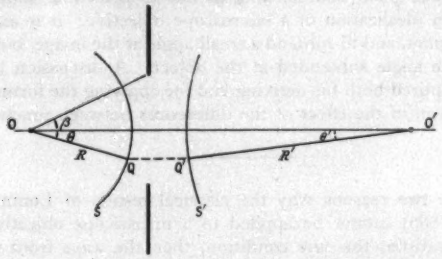
$$V(\mathrm{Q}) = V(\theta, \phi) = C \exp[ik \sin\theta(x\cos\phi + y\sin\phi)], \qquad \ldots\ldots(1)$$

approximately, provided that kr^2/R is small compared with unity, where $r^2 = x^2 + y^2$.

Now the condition imposed states that the resulting disturbance $V'(\mathrm{Q}')$ on S′ must be such as will converge to a focus at the point P′ whose coordinates with respect to O′ are $(-Mx, -My)$; if (R', θ', ϕ') are the polar coordinates with respect to O′ of the typical point Q′ of S′, then

$$V'(\theta', \phi') = F(\theta', \phi') \exp[i\alpha + ik'M \sin\theta'(x\cos\phi' + y\sin\phi')], \qquad \ldots\ldots(2)$$

where $F(\theta', \phi')$ is real, and α is a constant.



Out-of-focus diffraction patterns for microscope objectives.

Comparing equations (1) and (2), we see that the action of the lens may be thought of as transferring the disturbance from (θ, ϕ) on S to (θ', ϕ') on S′ with a constant change of phase, where

$$\phi' = \phi, \qquad \mu' M \sin\theta' = \mu \sin\theta. \qquad \ldots\ldots(3)$$

This thought may be expressed more rigorously as follows. The action of the lens may be represented by a Green's function:

$$V'(\theta', \phi') = \int_S G(\theta, \phi;\ \theta', \phi') V(\theta, \phi)\, d\sigma, \qquad \ldots\ldots(4)$$

where $d\sigma$ is the surface element of S. The range of integration should be the whole of the surface of S, but as we are going subsequently to impose an aperture, we shall consider only the contribution from the hemisphere $\theta \leqslant \pi/2$. Now the expressions $\exp[ik \sin\theta(x\cos\phi + y\sin\phi)]$ considered as functions of θ and ϕ, with kx, ky, as parameters, form a complete set of functions for the hemisphere —i.e. any function can be expressed as a linear combination of them, and they will be its (two-dimensional) Fourier components. But equations (1) and (2) show that for these functions an appropriate expression for $G(\theta, \phi;\ \theta', \phi')$ is

$$F(\theta', \phi') \exp(i\alpha) \,.\, \delta(\phi - \phi',\ \theta - s(\theta')), \qquad \ldots\ldots(5)$$

where $\delta(\psi, \chi)$ is a Dirac δ-function which is zero unless $\psi = 0$ and $\chi = 0$, and whose integral over the hemisphere is unity; and where $s(\theta')$ is given by

$$s(\theta') = \sin^{-1}(\mu' M \sin\theta'/\mu). \qquad \ldots\ldots(6)$$

It follows then that (5) is the unique function appropriate to the problem, and in conjunction with equation (4) gives expression to the idea of the preceding paragraph.

The amplitude factor $F(\theta', \phi')$ may be calculated by the following argument (Hopkins 1943). We suppose that all the energy radiated between θ and $\theta + d\theta$ converges between θ' and $\theta' + d\theta'$, where θ and θ' are related by equation (3), and therefore $\mu' M \cos\theta' \, d\theta' = \mu \cos\theta \, d\theta$; so that if A, A', are the amplitudes at corresponding points of S, S', then $A'^2 R'^2 \sin\theta' \, d\theta' = A^2 R^2 \sin\theta \, d\theta$, and so $\mu^2 A'^2 R'^2 \cos\theta = \mu'^2 A^2 R^2 M^2 \cos\theta'$, and hence

$$F(\theta', \phi') = A'/A = (\mu' MR/\mu R')(\cos\theta'/\cos s(\theta'))^{1/2}. \quad \ldots\ldots(7)$$

This completes the characterization of the ideal lens.

Consider now the disturbance on S due to an isotropic source at the point whose polar coordinates with respect to O are $(\rho, \eta, 0)$; it will be given approximately by

$$V(\theta, \phi) = A \exp[ik\rho(\cos\theta\cos\eta + \sin\theta\sin\eta\cos\phi)], \quad \ldots\ldots(8)$$

provided that $k\rho^2/R$ is small compared with unity. By equations (4), (5), (7), the corresponding disturbance on S' will be given by

$$V(\theta', \phi') = B(\cos\theta'/\cos s(\theta'))^{1/2} \exp[ik\rho\cos s(\theta')\cos\eta + ik'\rho M \sin\theta' \sin\eta\cos\phi'], \quad \ldots\ldots(9)$$

where the constant phase and amplitude factors have been absorbed in the constant B.

Now, since $k\cos s(\theta') \neq k' M \cos\theta'$, it will be seen that equation (9) does *not* represent a spherical wave front converging on the point which has coordinates $(M\rho, \eta, 0)$ with respect to O' unless $\eta = \pi/2$; and $\cos\eta(k\cos s(\theta') - k' M\cos\theta')$ may be taken as a measure of the spherical aberration which is introduced by the ideal lens when it is used to image a point which does not lie in the proper object plane.

Suppose that the light is restricted by a circular aperture which has its centre on the axis, lies on S', and whose radius subtends an angle β' at O'; then the disturbance at a point of the image plane whose coordinates are $(-Mx, -My)$ will be

$$D\int_0^{2\pi} d\phi' \int_0^{\beta'} \exp[ik\rho\cos s(\theta')\cos\eta + ik' M\sin\theta'\{(\rho\sin\eta - x)\cos\phi' - y\sin\phi'\}] [\cos\theta'/\cos s(\theta')]^{1/2}\sin\theta' \, d\theta'. \quad \ldots\ldots(10)$$

It is more convenient to express this as an integral over S, by making the substitution $\phi' = \phi$, $k' M\sin\theta' = k\sin\theta$; we also put $\cos\theta' = 1$ since we are assuming that β' is small. Further, x and y occur in the expression (10) only in the terms $(\rho\sin\eta - x)$, y, which give the x and y coordinates of the source *relative* to the point (x, y); there is therefore no loss of generality in taking $x = 0$, $y = 0$. Expression (10) now becomes

$$E\int_0^{2\pi} d\phi \int_0^{\beta} \exp[ik\rho(\cos\theta\cos\eta + \sin\theta\sin\eta\cos\phi)]\sin\theta\cos^{1/2}\theta \, d\theta, \quad \ldots\ldots(11)$$

where β is related to β' by

$$k\sin\beta = k' M\sin\beta';$$

so that β is the semi-angle of the aperture as seen from O. For an immersion lens of numerical aperture N, and oil of refractive index μ,

$$\sin\beta = N/\mu. \quad \ldots\ldots(12)$$

Expression (11) may further be simplified by putting

$$r=\rho\sin\eta, \qquad d=\rho\cos\eta,$$

and carrying out the integration with respect to ϕ:

$$H(r)=2\pi E\int_0^\beta \exp(ikd\cos\theta)J_0(kr\sin\theta)\sin\theta\cos^{1/2}\theta\,d\theta. \quad \ldots\ldots(13)$$

This equation gives the complex diffraction pattern $H(r)$ which arises when a point source which is a distance d out of focus is imaged by an ideal flat-field anastigmat of large magnification; r is the distance from the image of the source to the point of observation, measured in object distances (i.e. Mr is the actual distance).

If we take $E=3/[4\pi(1-\cos^{3/2}\beta)]$, then $H(0)=1$ when $d=0$.

Consider now a coherent line source, given by $z=d$, $y=c$; the amplitude of the resultant disturbance at O will be, by (11),

$$E\int_0^\beta \sin\theta\cos^{1/2}\theta\,d\theta\int_0^{2\pi}d\phi\int_{-\infty}^{+\infty}\exp[ik(d\cos\theta+c\sin\theta\sin\phi+x\sin\theta\cos\phi)]\,dx. \quad \ldots\ldots(14)$$

If the integration with respect to x is performed after the integrations with respect to θ and ϕ, then the result is convergent (though not absolutely convergent) when the limits of integration for x tend to infinity. It is, however, more convenient to evaluate the integrals in the order given by (14), and then some device—such as the use of a cut-off factor—must be applied to render the integral with respect to x convergent. A quick way of obtaining the answer (which can then be justified by rigorous analysis) is to use Dirac's δ-function, and the equation

$$\int_{-\infty}^{+\infty}\exp(ixu)\,dx=2\pi\,\delta(u). \quad \ldots\ldots(15)$$

Thus, putting

$$I=\int_0^{2\pi}\sin\theta\,d\phi\int_{-\infty}^{+\infty}\exp[ik(x\sin\theta\cos\phi+c\sin\theta\sin\phi)]\,dx,$$

we have

$$I=2\pi\int_0^{2\pi}\delta(k\sin\theta\cos\phi)\exp(ikc\sin\theta\sin\phi)\sin\theta\,d\phi. \quad \ldots\ldots(16)$$

Now put $u=k\sin\theta\cos\phi$, $du=-k\sin\theta\sin\phi\,d\phi$;

then $u=0$ when $\phi=\pi/2$, or $\phi=3\pi/2$, and so

$$\begin{aligned} I&=-2\pi\left\{\int_0^\pi+\int_\pi^{2\pi}\right\}\{(1/k\sin\phi)\exp(ikc\sin\theta\sin\phi)\,\delta(u)\,du\}\\ &=(2\pi/k)[\exp(ikc\sin\theta)+\exp(-ikc\sin\theta)]\\ &=(4\pi/k)\cos(kc\sin\theta). \end{aligned} \quad \ldots\ldots(17)$$

Using this result expression (14) becomes

$$L(c)=2F\int_0^\beta\cos^{1/2}\theta\cos(kc\sin\theta)\exp(ikd\cos\theta)\,d\theta. \quad \ldots\ldots(18)$$

This equation gives the diffraction pattern $L(c)$ due to a coherent line source which is a distance d out of focus; c is the distance of the point of observation from the line.

Equation (18) may also be written as

$$L(c)=F\int_{-\beta}^{\beta}\exp[ik(c\sin\theta+d\cos\theta)]\cos^{1/2}\theta\,d\theta; \quad \ldots\ldots(19)$$

this equation could also have been derived by considering a line source and an (infinitely long) *cylindrical* lens.

Discussion

We consider first the assumptions which have been made in deriving the formulae (13) and (18).

(A) The Kirchhoff assumption—that the presence of the lens and its surroundings do not affect the disturbance on S; in other words, no account has been taken of reflections from the lens and the microscope. It seems reasonable to suppose that reflection will be greatest (and hence transmission least) at the edges, and therefore a better approximation to the true state of affairs could perhaps be obtained by replacing the factor $\cos^{1/2}\theta$ in equations (13) and (18) by $\cos^{p}\theta$ with some value of p greater than $\frac{1}{2}$.

(B) The assumption that there is no diffraction within the lens—for the condition we have imposed means that the disturbance at a point of S′ is supposed to depend only on the value of the disturbance at the corresponding point of S. Since for a real lens the amplitude and phase at a point of S′ must depend on those for a *neighbourhood* of the corresponding point of S, it seems likely that in a real lens the interference effects which occur at the image will not be so sharp, and so the graphs of the patterns for a real lens may be expected to appear flatter than those calculated from equations (13) and (18). The departure of the real from the ideal will evidently become greater as the source is moved further out of focus, for the variation of phase and amplitude over S will then become larger and more rapid.

(C) The assumption that $(k\rho^2/R)$ is negligible. If $\rho = 10\lambda$, $R = 4000\lambda$ (corresponding to $R = 2$ mm, $\lambda = 5000$ Å) then the error in amplitude is certainly less than one twelfth of the amplitude at the centre of the in-focus pattern; it is probably a good deal less than this, but more accurate estimation is not easy.

(D) The assumption that the disturbance may be treated as a scalar quantity; this is necessary for the superposition of disturbances which is made in equation (10). The assumption is justified provided M is large, for then the angle of the converging wave front is small; in fact it can be shown that the error introduced by the assumption is of the order of $1/M$. Hopkins (1943) has discussed the effect of taking into account the vectorial nature of light.

It may also be pointed out that there is no simple solution of Maxwell's equations which corresponds to an *isotropic* source: the simplest solution corresponds to dipole radiation. A further reason for considering dipole sources is given below.

The results of this paper have already been used by Mitchison (1953) for the interpretation of photographs of objects whose dimensions are of the same order as the minimum distance of separation which can be resolved by the microscope; it is therefore relevant to consider what further assumptions must be made if the formulae are to be applied to actual microscopy. (Mitchison compared graphs obtained from photographs using a microphotometer with graphs of the calculated diffraction patterns of various hypothetical objects.)

The assumptions which must be made if the pattern due to a complex object is to be calculated are necessarily rather involved. For simplicity let us suppose that the object is illuminated from below by a plane wave W whose direction of propagation is parallel to the z axis. We divide the resultant field into two components: the incident field W, and the scattered field U. At any point of space we may regard U as a superposition of fields U_i, each U_i being the scattered field due to a small part P_i of the object considered. Then by a well-known

result of Lord Rayleigh's, if the dimensions of P_i are small compared with λ, U_i will be approximately the same as the field due to a dipole situated within P_i; the strength, direction, and phase of this equivalent dipole depend on the volume, the refractive indices, and the absorption coefficient of P_i, and on the field W_i which would exist in the neighbourhood of P_i if the material of P_i were removed. In certain circumstances W_i will be nearly the same as W.

The equivalent dipole at P_i may be resolved into a dipole a_i parallel to the z axis, and a dipole b_i perpendicular to it. The field at the image plane due to a_i will, presumably, be small, and will be neglected. The field radiated by b_i will be proportional to $\cos\theta$; so that, if P_i is at (x_i, y_i, z_i), the complex amplitude of the field at the point $(-Mx', -My')$ of the image plane will be

$$B_i \int_0^\beta \exp(ikz_i \cos\theta) J_0(kr \sin\theta) \sin\theta \cos^{3/2}\theta \, d\theta \qquad \ldots\ldots(20)$$

where B_i gives the amplitude and relative phase of b_i, and

$$r^2 = (x' - x_i)^2 + (y' - y_i)^2.$$

In general, the various dipoles b_i will not all be parallel, and so direct superposition of the diffraction patterns will not be permissible; but there are two special cases.

(1) If the material of the object is isotropic (i.e. not intrinsically birefringent), if its optical constants do not vary too rapidly (so that there is no form birefringence), and if the object is not too thick, then W_i may be equated with W, and the dipoles b_i will all be parallel to the electric vector of W. The image will be obtained by superposing the expressions (20), and the image due to W—which last will, presumably, be approximately uniform.

(2) If a polarizing microscope is being used, then only the components of the b_i parallel to the electric vector of the wave which is admitted by the analyser need be considered, and the corresponding patterns can be superposed.

So far we have considered only plane wave illumination; the effect of relaxing this restriction, i.e. of considering a condenser of finite aperture, will be twofold. Firstly, the fields scattered by different parts of the object will no longer be coherent. Secondly, the field scattered by a single part will be neither dipole nor isotropic radiation, but will lie somewhere between those two; and consequently the diffraction pattern will lie between the expressions (20) and (13). This suggests that

$$H(r) = G \int_0^\beta \exp(ikd \cos\theta) J_0(kr \sin\theta) \sin\theta \cos\theta \, d\theta \qquad \ldots\ldots(21)$$

should replace equation (13); and that

$$L(c) = G \int_0^\beta \exp(ikd \cos\theta) \cos(kc \sin\theta) \cos\theta \, d\theta \qquad \ldots\ldots(22)$$

should replace (18). Equation (21) reduces to the Airy disc formula when $d = 0$, and (22) to the formula of Rayleigh (1896): $L(c) = (\lambda/2\pi Nc) \sin(2\pi Nc/\lambda)$; and the integrals occurring in them are probably easier to compute than those which occur in equations (13) and (18). Hartree (1953) has devised methods for computing the integral in equation (22). The differences between the three formulae which have been given for the diffraction pattern of a point object —viz. expressions (13), (20), (21)—will of course only be appreciable when the

aperture of the objective is large; it is not easy to give precise estimates for these differences, and in practice they may be of small importance when compared with the discrepancies introduced by the departure of a given real lens system from the ideal objective here considered.

References

HARTREE, D. R., 1954, *Proc. Camb. Phil. Soc.*, **50**, 567.
HOPKINS, H. H., 1943, *Proc. Phys. Soc.*, **55**, 116.
LOMMEL, W., 1885, *Abh. Bayer. Akad. Wiss.*, **15**, 233.
MITCHISON, J. M., 1953, *J. Exp. Biol.*, **30**, 397.
RAYLEIGH, Lord, 1896, *Phil. Mag.*, **42**, 167.
STRUVE, H., 1886, *Mém. Acad. Sci. St. Pétersb.* (*Sci. math., phys., nat.*) (7), **34**, No. 5, 1.

Mechanisms of Creep in a Precipitation Hardened Alloy†

By G. C. E. OLDS
Research Laboratory, The British Thomson-Houston Company Limited, Rugby

M.S. received 20th May 1954, and in amended form 9th August 1954

Abstract. The binary alloy 97% copper–3% silver was selected as the most ideal alloy for research on the creep process in a precipitation hardened alloy. A heat treatment was divised which gave a slightly overaged Widmanstätten structure of high internal hardness and maximum stability. Wires of the alloy were subjected to deformation over a range of constant stresses at temperatures from 100°C to 300°C. Two mechanisms of transient creep were observed and were found to be complementary. The first mechanism occurred at the lower temperatures giving creep strain proportional to ln (time), whilst the second mechanism in which creep was proportional to $(\text{time})^{1/3}$ appeared and became dominant at higher temperatures. A quasi-viscous creep mechanism was also observed which at high temperatures was more important than the transient process. The linear nature of the quasi-viscous creep was limited by a critical time, and was believed to be governed by the plastic deformation which occurred on loading. The three mechanisms observed agree with some theoretical analyses which have been put forward.

§ 1. Introduction

Systematic study of the creep of metals has been up to now mainly confined to pure metals, owing to the complexity of the fundamental mechanisms involved. The present work extends the systematic examination of the creep curve to a simple binary alloy in which the crystal lattice of one metal is hardened by precipitated particles of a second metal. In the many creep experiments which have been performed on commercial precipitation hardened alloys, the complexity of the structures and their instability during the experiments have precluded conclusions being drawn on the fundamental laws of creep in a hardened alloy. The particular alloy used in the present work was chosen to avoid structural instability during experiment, and to enable sensitive measurement of creep to be made over a range of temperatures and stresses. The form of the creep curve observed under these conditions is used as a basis for analysis of the deformation mechanisms; a summary of the results obtained has appeared elsewhere (Olds 1953).

In pure polycrystalline metals, the creep strain which occurs at constant stress and temperature has been shown (Andrade 1910, 1914; Cottrell and Aytekin 1950) to be of the form

$$\epsilon = \beta t^{1/3} + \gamma t + \delta, \qquad \ldots\ldots(1)$$

where δ is the instantaneous plastic strain on loading, and β and γ are constants governing transient and steady-state or quasi-viscous creep respectively. The

† This paper is a modified summary of work for which the degree of Doctor of Philosophy was awarded by the University of London.

available evidence suggests that γ depends exponentially on stress and temperature.

$$\gamma=\gamma_0 \exp(b\sigma), \qquad \ldots\ldots(2)$$

$$\gamma=C\exp(-H/RT). \qquad \ldots\ldots(3)$$

At low temperatures γ is negligible and only transient creep occurs. Recent work on pure polycrystalline copper, within this low temperature range (Wyatt 1953), provides evidence that transient creep can itself be divided into two components. The β term occurs at the higher temperatures, but at much lower temperatures this is entirely replaced by the other transient component:

$$\epsilon=\alpha \ln t+\delta. \qquad \ldots\ldots(4)$$

At intermediate temperatures, transient creep is represented by a combination of the α and β terms.

Little work has so far been carried out on precipitation hardened alloys which yields results enabling comparison with pure metals. Davis and Thompson (1950) have performed low temperature creep tests on a copper–3% silver alloy and observed transient creep strain proportional to $(\ln \nu t)^{2/3}$. They concluded that the behaviour of this alloy was quite different from that of many pure metals which gave curves in accordance with equation (1).

§ 2. Selection of Alloy

For the present work, it was desirable to select a simple precipitation hardened alloy with structure furthest removed from that of a pure metal, in that high internal hardness imparted great resistance to creep deformation. Furthermore, if the creep curves obtained were to be analysed and compared, it was necessary for all experiments to be carried out on an identical and stable metallurgical structure.

A survey of binary equilibrium diagrams suggested that copper base silver alloy (figure 1) most nearly fulfilled the ideal conditions, since almost complete precipitation of silver was possible above 300°c, leaving a considerable range below this temperature in which there would be little or no phase transformation. By slowly cooling copper–3% silver alloy (Barrett, Kaiser and Mehl 1935), it is possible to obtain the silver precipitate in the form of a Widmanstätten structure. This implies continuous precipitation in which there would be no grain fragmentation (Jones, Leech and Sykes 1942, Geisler 1951). The composition 97% copper–3% silver was used for the present work with 300°c as the upper limit of temperature for creep testing. Stability of structure below 300°c was aided by the very slow diffusion of silver in copper.

§ 3. Heat Treatment of Wires

The alloy used was of spectroscopic purity with uniform silver content equal to 2·94% by weight. It was supplied as 16 s.w.g. (0·064 in. dia.) cold drawn wire of hardness 160 (Vickers diamond pyramid scale). Trial heat treatments were made to obtain the alloy in a hardened and stable condition. The first stage consisted of a high temperature anneal at 780°c (figure 1) to allow complete solution and homogenization of the silver content. This was followed by a cooling treatment from A to B which is at 300°c. Precipitation took place between C and B. The alloy was then held at 300°c so that the precipitation process was completed, and over-ageing occurred. An atmosphere of nitrogen

was used for these operations, and temperature uniformity maintained to within 1°c.

Precipitation between 780°c and 300°c was tried at various cooling rates, and it was found that both considerable hardening and the desired Widmanstätten structure occurred only when very small cooling rates were used. The heat treatment schedule finally adopted for the wires consisted firstly of the solution treatment at 780°c for 70 hours, followed by a 5°c/hour cooling treatment from 780°c to 300°c. This produced an average hardness of 75, compared with 58 for water-quenched specimens. The precipitation was followed by an overaging treatment of 90 hours at 300°c to stabilize the alloy. This reduced the hardness to 73. The specimens were then given a further 200 hours stabilizing treatment at the temperature of the subsequent creep experiments. This was an extra safeguard against instability and had negligible effect on the hardness. It was thus possible to provide a stable and almost identical metallurgical structure which could be examined by creep deformation over a range of temperature up to 300°c. The Widmanstätten structure obtained is shown in figure 9 (Plate), and consisted of a dense array of silver-rich plates across each grain, somewhat spheroidized in the vicinity of the grain boundary, with larger globules of silver in the boundary.

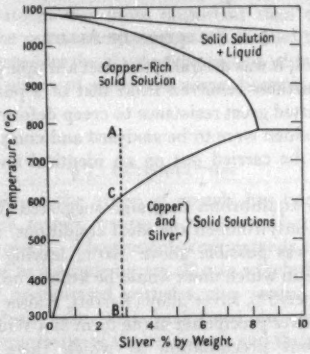


Figure 1. Copper–silver constitutional diagram at copper-rich end.

§ 4. Creep Apparatus

The wire specimen, of effective length 4 in., was held at each end by split steel collets in an upper and lower grip to which the arms of a Lamb's extensometer were attached. A pair of knife-edges, crossed at 90°, was used at each end to ensure axial alignment of stress. The means by which constant stress was applied is illustrated in figure 2. As the creep deformations measured were very small, never exceeding 0·8%, it was possible to use a beam of very simple design, based on a method employed by Carreker (1950). The load was applied at the extremity of the arm A which was bolted at right angles to the horizontal part B. The arm C was present to obtain symmetry and balance about the knife-edge D. The condition for constant stress in the specimen, assuming the cross sectional area of the latter decreases uniformly along its length, is

$$L/L_1 = L_2/L_3, \qquad \ldots\ldots(5)$$

where L is the length of the specimen when the angle (θ) between the beam and horizontal is zero. Stress was transmitted to the specimen by the two steel wires E and F which were respectively linked with the knife-edge G and the ball seating H. The screw J enabled the scale pan to be loaded without transmitting stress to the specimen column.

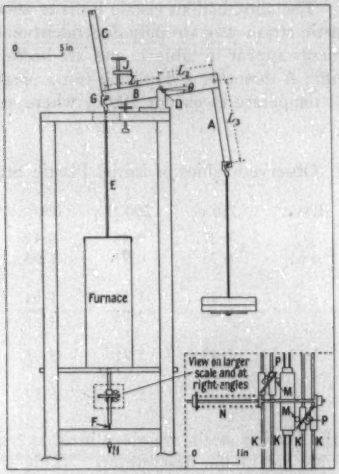


Figure 2. Diagram of creep apparatus.

Axial creep strain in the specimen was transferred by light steel arms K to rotation by two rollers P to which plane mirrors were fitted. A second pair of rollers M separated the inner extensometer arms from the central column. The springs N applied a small pressure to the rollers to provide sufficient frictional contact. Readings were taken to 1×10^{-6} strain by telescopic viewing of a distant illuminated scale from which light was reflected at both mirrors.

The furnace used had three independent windings so that temperature adjustments along the specimen could be made. Measurements of temperature were taken to within 0·1°C, and control of constant and uniform specimen temperature was obtained to $\pm 0{\cdot}2$°C throughout the experiments.

§ 5. Creep Experiments

The basis of the experiments was to submit the alloy wires to constant uniaxial stresses at constant temperatures, and observe the plastic deformation which occurred as a function of time. Comparatively small total deformations were used so that general stability of the structure was maintained. The grain size given by the standard heat treatment was 110–120 grains mm^{-2}.

The initial setting of θ was based on the expected instantaneous plastic strain on loading so that the beam was very nearly horizontal after the full stress had

been applied. Rotation of the screw J enabled stress to be applied at approximately uniform strain-rate over a period of 15–30 seconds. Strain readings were taken at least every two minutes up to 350–400 minutes, and in some cases at further intervals up to 1500 or 2700 minutes.

The temperatures at which tests were carried out were 100°C, 150°C, 200°C, 250°C, 275°C, and 300°C. The range of stresses employed was based on the results of trial tests. The alloy had an elastic limit at about 9 km mm^{-2}, above which the initial plastic strain was strongly dependent on stress. The stresses used for the experiments appear in table 1, and are hereafter designated by the numbers 1–6, the suffix A denoting a duplicate test. Tests were performed at the six stresses at all temperatures except 100°C, where stresses 2 and 5 only were used.

Table 1. Observed values of Initial Plastic Strain δ (%)

Stress (kg mm^{-2})		100°C	150°C	200°C	250°C	275°C	300°C
1	9·25		0·37	0·48	0·42	0·42	0·52
2	10·50	0·61	0·73	0·91	1·05	1·10	1·32
2_A							0·93
3	11·75		1·25	1·51	1·61	1·58	1·71
4	13·00		1·99	2·75	2·77	2·84	3·59
4_A							2·81
5	14·25	2·42	3·48	3·71	3·95	4·03	4·42
5_A						4·21	4·54
6	15·50		4·96	5·31	5·57	5·82	6·25

Values of the initial loading strain are given in table 1 and refer to the total plastic strain up to the end of the first minute. This is because creep was initially very rapid and extrapolation to zero time could not be carried out. Although the table shows some inconsistency between different wires, graphs of log (initial plastic strain) against T^{-1} at constant stress showed approximate linearity suggesting the relationship

$$\delta = \delta_0 \exp(-J/RT). \qquad \ldots\ldots(6)$$

δ is the initial plastic strain at T°K and J, the activation energy of plastic strain per gramme-atom, gave an approximately constant value of 900–1200 calories.

The time dependent deformations observed up to 350–400 minutes at the various temperatures are illustrated by figures 3 and 4, which refer to tests at 150°C and 275°C respectively. Since the total creep strains measured were much smaller than the plastic strains on loading, the latter were omitted on these graphs and the initial ordinate for each curve chosen arbitrarily, so that a given family of curves could be represented in a single diagram.

A thin oxide film was formed during the experiments on wires tested at 200°C and above, but the results do not suggest that this had any significant effect on deformation. Microscopical examination of the etched wire sections after testing showed no observable change in structure.

§ 6. Analysis of Creep Curves

The technique used is to observe the mathematical form of the curves obtained at different temperatures, and to discuss the results in terms of current theories relating creep strain, time, temperature and stress.

6.1. 100°C *and* 150°C *Curves*

A comparatively high creep-rate at the beginning of these tests falls rapidly to a small value. A graph of creep strain against log (time) shows good linearity giving the equation

$$\epsilon = \alpha \ln t + \delta. \qquad \ldots\ldots(7)$$

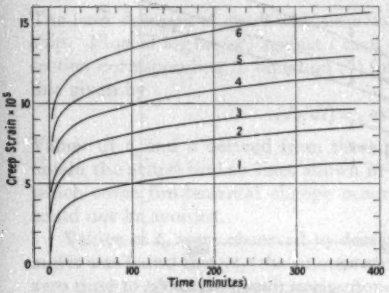


Figure 3. Creep curves for 150°C.

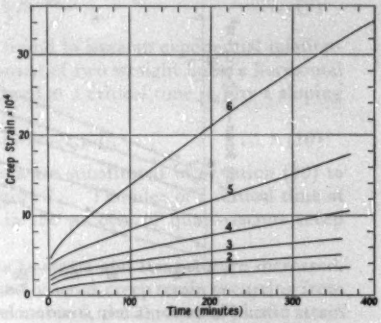


Figure 4. Creep curves for 275°C.

The plots for the 100°C curves are shown in figure 5. For the 150°C curves, the plots indicate that a better linear fit for long times is obtained if a very small constant creep-rate term is added to the right-hand side of equation (7). The magnitude of this term increases from 10^{-9}/minute for low stresses to 10^{-8}/minute for high stresses.

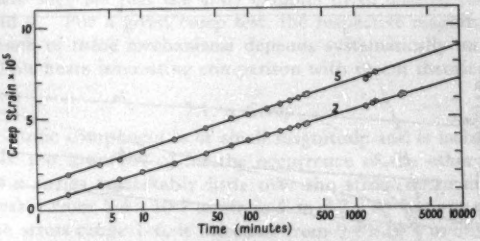


Figure 5. Creep strain against ln (time) at 100°C.

6.2. 200°C *and* 250°C *Curves*

Creep-rate at these temperatures falls off less rapidly with time, and it was found that the main component of transient creep obeyed the $t^{1/3}$ law. A small logarithmic component was also evident at the beginning of each test. At the two highest stresses at 200°C, a small quasi-viscous creep term could be distinguished. The latter was much larger at 250°C. The creep curves observed up to 250°C can therefore be accounted for by the equation

$$\epsilon = \alpha \ln t + \beta t^{1/3} + \gamma t + \delta. \qquad \ldots\ldots(8)$$

Figure 6 shows a $t^{1/3}$ plot for the 200°c curves in which the β term is dominant over the α and γ terms. Non-linearity shown at small times suggests that α is not constant with stress as has been assumed in this plot.

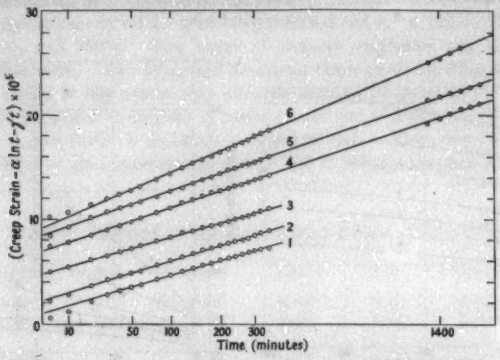


Figure 6. (Creep strain$-\alpha \ln t-\gamma t$) against (time)$^{1/3}$ at 200°c.

6.3. 275°c *and* 300°c *Curves*

The γ component becomes dominant at these temperatures, and the α term, if present, is negligible. Plots of $d\epsilon/dt$ against $t^{-2/3}$ are linear in accordance with equation (8) ($\alpha=0$) for the lower stresses at 275°c, but for higher stresses at 275°c and all stresses at 300°c a new phenomenon occurs at longer times by which

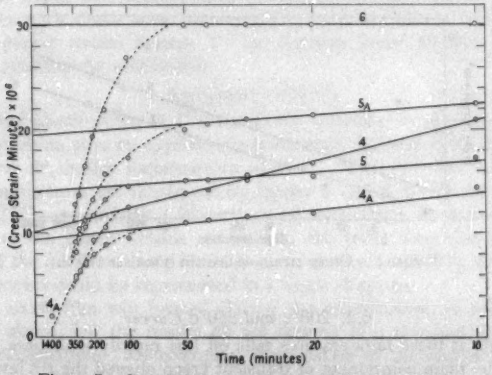


Figure 7. Creep strain-rate against (time)$^{-2/3}$ at 300°c.

$d\epsilon/dt$ decreases below its expected value. This is illustrated by figure 7, for stresses 4–6 at 300°c, where the effect is most marked. The straight lines correspond to equation (8) whilst the broken lines show the observed deviation from this.

The new effect represents some mechanism which affects the quasi-viscous creep-rate (γ), and reduces it steadily with time. This could not be connected with the smaller β term, and for the purposes of analysis, the latter is subtracted from the observed creep strain to give a new strain variable η. Thus, while equation (8) is obeyed, $d\eta/dt$ is a constant:

$$\frac{d\eta}{dt} = \frac{d\epsilon}{dt} - \frac{1}{3}\beta t^{-2/3} = \gamma. \qquad \ldots\ldots(9)$$

The later decrease of $d\eta/dt$ with time was found to have an exponential relationship. Plots of $\log(d\eta/dt)$ against t each consist of two straight lines, a horizontal section corresponding to equation (9) followed at a critical time (t_0) by a sloping line given by

$$(d\eta/dt) = \gamma \exp\{-(t-t_0)/a\}. \qquad \ldots\ldots(10)$$

Values of t_0 and a derived from these plots are substituted in equation (10) to obtain the actual broken lines shown in figure 7. The idea of a critical time at which some fundamental change occurs in the process of quasi-viscous creep could not be avoided.

Values of t_0 were observed to decrease as stress and temperature increased, and it was found that γt_0 (the amount of quasi-viscous creep strain occurring from zero time to t_0) was approximately proportional to δ, the amount of plastic strain which occurred on loading. The ratio $\gamma t_0/\delta$ is approximately 0·03% creep : 1% initial plastic strain for all tests at 300°C and 275°C where t_0 was observed. This relationship does not lack generality with tests conducted at lower stresses at 275°C, and at lower temperatures, since, if t_0 is calculated as $0{\cdot}03\delta/\gamma$, it is found to exceed the maximum time of these experiments.

§ 7. Discussion

The analysis suggests that the alloy exhibits three distinct mechanisms of creep, α, β and γ. For a given creep test, the respective magnitudes of creep strain due to each of these mechanisms depends systematically on temperature and stress. This bears interesting comparison with recent theories of creep.

7.1. α *Creep*

The logarithmic component is of small magnitude and is isolated when the temperature is too low (100°C) for the occurrence of the other mechanisms. The constant α varies remarkably little over the stress range employed. At 100°C, α increases from $7{\cdot}4\times10^{-6}$ at stress 2 to $8{\cdot}2\times10^{-6}$ at stress 5, whilst at 150°C over the stress range 1–6, it increases from $9{\cdot}4\times10^{-6}$ to only $12{\cdot}1\times10^{-6}$. At higher temperatures α is reduced in the presence of β creep, and at 200°C–250°C it can be taken as a constant of value $6{\cdot}5\times10^{-6}$. At 275°C and above, the α component is negligible.

The existence in this alloy of two components of transient creep, α and β, bears close similarity with the results of Wyatt (1953) on pure metals. Wyatt found little variation of α with stress in pure polycrystalline copper, but its value at 100°C–150°C was about 200 times larger than for the present alloy under similar stresses. The simple logarithmic equation (7) has been put forward theoretically by Smith (1948) and by Cottrell (1952) whose treatment is based on the dislocation theory of 'exhaustion hardening', the obstacles to slip being dislocations in planes other than the slip plane. In the copper–silver alloy, the

obstacles to dislocation movement must be the strains surrounding precipitate particles. For such an alloy Mott and Nabarro (1948) and Mott (1953) have put forward the $(\ln vt)^{2/3}$ equation which is supported by the work of Davis and Thompson (1950) already referred to. Support for this equation was not found in the present work in which, however, the alloy used was considerably harder and of different structure from that used by Davis and Thompson.

7.2. β *Creep*

The observed values of β appear in table 2. The component is absent below 200°c, but above this temperature shows a general increase in magnitude up to the maximum temperature used. This suggests that the β term is a second distinct mechanism of transient creep which is dominant at higher temperatures than the

Table 2. Value of Constant $\beta \times 10^5$ (minutes)$^{-1/3}$

Stress	200°c	250°c	275°c	300°c
1	1·16	2·02	4·97	6·97
2	1·21	2·54	5·33	7·77
2_A				9·38
3	1·14	4·33	5·61	6·33
4	1·35	4·53	4·97	14·90
4_A				6·60
5	1·39	3·92	4·35	4·30
5_A			5·49	4·50
6	1·71	4·98	8·52	0

α mechanism. The constant shows a steady increase with stress at constant temperature, except at 300°c, where it passes through a maximum and falls to zero at stress 6. The latter phenomenon, which was confirmed by duplicate tests, is believed to be due to the high initial plastic strains (table 1) using up some or all of the available slip processes which would otherwise be activated as transient creep. It is known (Andrade 1952) that cold working a metal, prior to stressing in the same direction in a creep test, tends to eliminate the transient component of creep.

The β component of creep in this alloy is believed to correspond to the similar component commonly observed in pure metals by Andrade and later workers. The minimum temperature required for the occurrence of this mechanism suggests that crystal recovery is a necessary part of the process. On comparing with Wyatt's results for pure copper, in which both α and β creep were observed, the β term in the present alloy first occurs at least 100°c higher than in pure copper. This appears to be in agreement with the much higher recovery temperature which the alloy would exhibit. The magnitude of β for the alloy is also much reduced. When the temperature of the copper–silver alloy is 100°c higher than the pure metal, β for the alloy is still 100 times smaller than its value for pure copper.

The α component then appears to occur when hardening takes place without recovery. In this sense, it can be regarded as complementary to the β component.

7.3. γ *Creep*

The γ term represents constant rate quasi-viscous creep which can be considered independently of the two transient terms. The observed values of γ (strain/minute) appear in figure 8, in which a fairly linear relationship is

obtained by plotting $\log\gamma$ against stress at constant temperature. This confirms the equation

$$\gamma=\gamma_0\exp(b\sigma) \qquad \ldots\ldots(11)$$

which has been found to apply to pure metals. The curves give b as increasing from 0·6 to 0·7 $(\text{kg mm}^{-2})^{-1}$ with temperature, and γ_0 increasing rapidly with temperature. At constant stress, $\log\gamma$ varies linearly with T^{-1} in conformity with the equation

$$\gamma=C\exp(-H/RT). \qquad \ldots\ldots(12)$$

The activation energy H has a mean value of 42 000 calories/gramme-atom, which is a typical value for quasi-viscous creep in polycrystalline metals, and also for viscous flow in amorphous materials (Zener 1948).

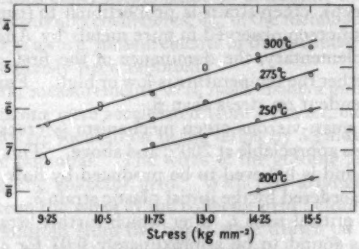


Figure 8. $\log\gamma$ plotted against stress.

The above evidence suggests that the γ mechanism is of similar origin to the linear component of creep in pure metals, and therefore related to boundary creep of a quasi-viscous nature. The alloy, however, exhibits a new feature at large values of γ in that the linear nature of this component is limited by a critical time, beyond which strain hardening occurs (equation (10)). In these circumstances, γ cannot be regarded as a genuine steady state creep-rate of the type which occurs in pure metals. The behaviour of the alloy, in regard to the critical time and the constancy of the ratio $\gamma t_0/\delta$ can be accounted for by the following hypothesis. The period of quasi-viscous creep results from a disordered boundary structure produced by the initial plastic strain on loading. The disorder takes the form of atoms, at grain boundaries and perhaps potential sub-boundaries, which, owing to the catastrophic nature of dislocation movements, have become unattached to any crystal lattice. The amount of disorder produced is proportional to the initial plastic strain δ. As the creep experiment begins, viscous-like flow can occur in the disordered regions until the harder crystals interlock. The effective shear stress across a boundary is constant during this period, and a constant macroscopic creep-rate is observed. The limiting creep strain at which the grains begin to interlock is γt_0 which is proportional to the initial amount of disorder, and therefore to δ. The function ascribed to crystal boundaries by this hypothesis can be supported by the experimental results of Kê (1947) and Andrade and Jolliffe (1952).

The strain hardening which follows the critical time closely obeys the exponential equation (10) (figure 7). At much larger times, however, there is a divergence from this relationship suggesting a recovery effect tending to balance the strain hardening. This is illustrated by test 4A (figure 7) in which the

experimental points lie close to the theoretical curves up to 350 minutes, but the value of $d\eta/dt$ at 1400 minutes is over twice the theoretical value. At very long times, it is believed that this leads to a small constant creep-rate which is the true steady state condition of creep.

§ 8. Conclusions

Copper–3% silver alloy can be precipitation hardened to give a stable structure which has very high resistance to creep deformation. There are two distinct transient creep processes which take place, one occurring mainly at the lower temperatures (100°C and 150°C), the other appearing at 200°C and becoming dominant above this temperature. In the former (α creep), creep strain is proportional to ln(time), and represents crystallographic hardening without recovery. In the latter (β creep), creep strain is proportional to $(\text{time})^{1/3}$ and is of similar form to the transient creep observed in pure metals by Andrade. The α and β processes are complementary, the dominance of the first or second depending respectively on whether the temperature is low or high. For a given temperature, α is much less dependent on stress than β.

There is also a quasi-viscous creep mechanism (γ creep) associated with the alloy, which becomes appreciable at 200°C and above. This creep strain increases linearly with time, and is believed to be produced by flow at crystal boundaries which have been disordered by the initial plastic strain δ. The linearity of creep exists only up to a critical time t_0, after which further strain-hardening occurs. The ratio $\gamma t_0/\delta$ was found to be approximately 0·03 for all experiments where t_0 occurred within the period of the test. The constant γ varies regularly with stress at constant temperature according to the equation $\gamma=\gamma_0 \exp(b\sigma)$. At constant stress, the variation of γ with temperature is given by $\gamma=C\exp(-H/RT)$.

Acknowledgments

The work described in this paper was carried out at the Research Laboratory of the British Thomson-Houston Co. Ltd., Rugby, Warwickshire. The author wishes to thank Mr. J. V. Hardwick for discussions on the subject matter, and Mr. L. J. Davies, Director of Research for permission to publish the paper.

References

Andrade, E. N. da C., 1910, *Proc. Roy. Soc.* A, **84,** 1; 1914, *Ibid.*, **90,** 329; 1952, *J. Iron and Steel Inst.*, **171,** 217.

Andrade, E. N. da C., and Jolliffe, K. H., 1952, *Proc. Roy. Soc.* A, **213,** 3.

Barrett, C. S., Kaiser, H. F., and Mehl, R. F., 1935, *Trans. A.I.M.E.*, **117,** 39.

Carreker, R. P., 1950, *J. Appl. Phys.*, **21,** 1289.

Cottrell, A. H., and Aytekin, V., 1950, *J. Inst. Met.*, **77,** 389.

Cottrell, A. H., 1952, *J. Mech. Phys. Solids*, **1,** 53.

Davis, M., and Thompson, N., 1950, *Proc. Phys. Soc.* B, **63,** 847.

Geisler, A. H., 1951, *Symposium on Phase Transformations in Solids* (London : Chapman and Hall).

Jones, F. W., Leech, P., and Sykes, C., 1942, *Proc. Roy. Soc.* A, **181,** 154.

Kê, T. S., 1947, *Phys. Rev.*, **71,** 533.

Mott, N. F., 1953, *Phil. Mag.*, **44,** 742.

Mott, N. F., and Nabarro, F. R. N., 1948, *Conference on Strength of Solids* (London : Physical Society).

Olds, G. C. E., 1953, *Nature, Lond.*, **172,** 999.

Smith, C. L., 1948, *Proc. Phys. Soc.*, **61,** 201.

Wyatt, O. H., 1953, *Proc. Phys. Soc.* B, **66,** 459.

Zener, C., 1948, *Elasticity and Anelasticity of Metals* (Chicago : University Press).

RESEARCH NOTES

The Secondary Electron Emission of Sodium and Zinc

By J. WOODS
Research Laboratories of the General Electric Co. Ltd., Wembley, Middlesex

MS. received 6th July 1954

THIS paper describes experiments made to determine the maximum values of the secondary emission coefficients, δ_m, of the metals sodium and zinc. As far as the author is aware, measurements of the secondary emission of these elements have not previously been reported.

In the present work, measurements were made using a guided grid secondary emission valve. Electrons were accelerated from a planar, oxide-coated cathode by a control grid, and were focused on to the target by a guiding grid, which was maintained at cathode potential. The wires of the guiding grid were accurately aligned with those of the collector grid. With this arrangement the primary electrons passed between the wires of the collector grid before striking the target. Secondary electrons emitted from the target were drawn off to the collector by the potential difference applied between those electrodes. The target was a large copper block which could be water cooled if required.

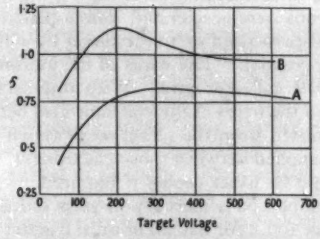
The valves were pumped on a mercury-in-glass diffusion pump with liquid air trap. After baking at 400°C for half an hour, the getter was outgassed and the cathode was heated at 1050°C for two minutes to break down the carbonate coating to oxide. The getter was then dispersed and the valve sealed off from the pump. After activating the cathode the metal to be examined was deposited on the copper target. A primary current density of 20 μA cm^{-2} was used to bombard the target, and the collector potential was 150 V positive with respect to that of the target. This was more than sufficient to saturate the secondary current. The interception of the primary beam by the collector was of the order 1–2%.

The accuracy of the valve system was determined by carrying out some preliminary experiments on evaporated films of barium, since the secondary emitting properties of this metal are well known. Two batalum getter strips were used as sources of barium, and they were positioned so that barium could be distilled on to the target without fouling the other electrodes. A value of 0·91 for δ_m at a target potential of 350 V was found, which agrees with the result of Bruining and de Boer (1938).

In order to introduce sodium on to the copper block, the electrolytic method developed by Forro and Patai (1931) was used. In these experiments soda-glass was used for the valve envelope, and an auxiliary tungsten filament was mounted on the pinch. The valves were pumped in the normal way, and after seal off were partially immersed in a bath of molten sodium nitrate which melts at 310°C. The valve was positioned in such a way that the glass opposite the tungsten filament was immersed in the electrolyte, while the target was water cooled. With the filament at 2500°K a potential of 240 V was applied between it and a copper anode immersed in the nitrate. Under these conditions a current of about 5mA

was drawn from the filament and sodium was deposited on the target. The mechanism of this process is as follows. Sodium ions pass through the electrolyte under the action of the applied field and reach the glass envelope. Since the nitrate and the glass in contact with it are at 310°C, the glass is also an electrolytic conductor, and therefore the sodium ions pass through the glass. On the inside of the envelope the ions are neutralized by the electrons from the filament, and the tube is filled with sodium vapour, which then condenses on the water-cooled target.

At various stages during this process the valve was removed from the molten nitrate, and the secondary emission of the target was measured. During the first determinations the copper colour of the target block could be seen through the sodium film which had a milky appearance. For these films δ_m was about 1·2. However, after considerably more sodium had been deposited on the target, so that no copper colour could be seen at all, δ_m fell to a constant value of 0·82. This maximum value was reached at a target potential of 300 v as curve A in the figure shows.



Secondary electron emission from : A, sodium and B, zinc.

Evaporated films of zinc were formed on the targets of sealed off valves in the following way. A small piece of zinc was placed in a molybdenum evaporator cup which was equipped with a tungsten wire heater. Because the evaporator assembly was comparatively bulky, the cathode and grids of the valve were assembled in such a way that they could be slid out of position facing the target while zinc was evaporated on to it. Subsequently they were returned to their original position and measurements were made. During this work it was found that if the copper was at room temperature while the zinc was distilled, hardly any zinc was condensed on the target. Instead it settled on the glass at the opposite end of the valve facing the target and behind the evaporator. This was due to the small accommodation coefficient of copper for zinc at room temperature. To overcome this difficulty liquid air was placed in an external cavity in the copper block while the zinc was evaporated, then a thick metallic coating was formed on the target. Curve B in the figure shows the variation of δ_m with target voltage for a thick coating of zinc. The maximum value of δ was 1·15 at a target potential of 200 v.

References

Bruining, H., and de Boer, J. H., 1938, *Physica*, **5**, 17.
Forro, M., and Patai, I., 1931, *Z. Tech. Phys.*, **12**, 256.

An Apparatus for Heat-Treatment and Quenching Small Specimens in a Vacuum from High Temperatures

By T. H. SCHOFIELD
Metallurgy Division, National Physical Laboratory, Teddington, Middlesex

Communicated by N. P. Allen; MS. received 28th July 1954

§ 1. Introduction

Several methods of heating and quenching metal specimens from temperatures up to about 1600°c are described in the literature. Such furnaces usually consist of a refractory tube wound with platinum or molybdenum tape or wire and a mechanical or electrical device for dropping the specimen into a quenching bath placed either within or external to the furnace. In either case it is usually necessary to fill the furnace tube with inert gas before quenching. Schramm, Gordon and Kaufmann (1950) describe an induction furnace which can be used for heat-treatment and quenching specimens from temperatures up to 2400°c. The specimen is suspended on a tungsten wire inside a tungsten heating chamber which is enclosed in and supported on zirconia tubes. Quenching is effected by loosening a clamp, which supports one end of the tungsten suspension, by external manipulation thus allowing the specimen to fall into a bath of silicone oil remote from the hot zone. The temperature of the specimen is measured optically.

§ 2. Furnace

The furnace and quenching apparatus are shown in the heating position (figure 1†) and at the instant of quenching (figure 2†). The furnace is similar to that described by Alberman (1950) and has also been used by the author for melting point determinations (Schofield and Bacon 1954). The heating element consists of a spiral of tungsten wire enclosed in a series of radiation shields. The base screen E (figure 2) rests on three screws, placed 120° apart, which project through the copper block F (figure 2) attached to one of the water cooled electrodes. The lower end of the heating element projects through a hole in the base radiation shield and is attached to the copper block by means of a set screw. In the original furnace (Alberman 1950, Schofield and Bacon 1954) the upper end of the heating element projected through a hole in the top radiation shield before connection to the electrode but in order that this shield could be raised when the furnace was hot, the upper end of the element is bent at right angles and passes through a hole in the cylindrical (side) radiation shield, from which it is insulated by a short length of sintered alumina tube, before connection to the electrode. This insulation is necessary otherwise some current will flow through the radiation shields.

§ 3. Crucible

The specimen is contained in a cylindrical molybdenum crucible A, 5/8 in. long, 5/16 in. outside diameter and 1/32 in. in wall thickness (figure 2), which is closed with a loosely fitting molybdenum stopper drilled with an axial hole 1/16 in. diameter. When the crucible is heated to a uniform temperature, the temperature as determined by sighting on the hole by means of an optical pyrometer approximates to that of a full radiator (black body). The size of the specimen is not

† See Plate.

critical provided that its upper surface is at least 5/16 in. from the hole in the stopper so as to ensure that the observed temperature does not depart appreciably from that of a full radiator (Buckley 1934) and that it does not touch the crucible walls. Further to minimize risk of contamination the specimen rests on a small quantity of coarsely powdered material of the same composition. In practice the size of the specimen was about 3/16 in. cube which in the case of titanium-rich alloys weighs about 0·5 g. A check on the accuracy of the temperature measurements was carried out using two small cylindrical specimens of Van Arkel titanium, each 1/8 in. diameter and 1/8 in. long. The base of each specimen was recessed leaving a rim 1/64 in. wide and deep. The specimens rested on a thin disc of zirconia placed at the bottom of the molybdenum crucible and were slowly heated in turn to 1660° and 1670°C and held at these temperatures for two minutes. Visual observation of the specimens after cooling showed that melting had occurred at 1670°C but not at 1660°C indicating a melting point between these temperatures which agrees with the previous determinations (Schofield and Bacon 1954). The molybdenum crucible has been used for work on the constitution of the titanium–oxygen system but for the investigation of other systems it may be preferable to use crucibles of different metals or refractory materials.

§ 4. Quenching Device

The following description of the quenching device refers particularly to figure 2. The molybdenum crucible A rests on a refractory (zirconia or thoria) disc B which in turn rests on a molybdenum platform C, all of which are normally inside the heating coil. The molybdenum platform is supported on a sintered alumina tube D (figures 1 and 2) which passes through a central hole in the lower radiation shield E and a hole in the copper block F attached to one of the electrodes. The lower end of the alumina tube rests in a central hole in a steel bush G (figures 1 and 2). The lower end of the steel bush is located in a brass rod H which is led through an O ring seal below the furnace base. This rod is supported in a brass guide screwed to the underside of the furnace base and can be raised by manual operation of a lever attached to a cam (not shown in figures 1 and 2). Full movement of the cam raises the molybdenum platform on which the crucible rests about 3½ in. which ensures that the crucible is well outside the heating coil and shields. A steel arm I is rigidly attached to the bush G; the top of this arm is connected to the top radiation shields by means of a steel side arm J and molybdenum wire stirrup K. The stirrup is insulated from the radiation shields by sintered alumina tubes, which pass through tantalum strips welded on to the shields. An adjustable steel cam L is also fixed to the side arm I. The upper end of a vertical steel rod O screwed to the furnace base supports two horizontal arms, about 40° apart, one of steel M and the other of molybdenum N (figures 1 and 2). Both arms are free to rotate in a horizontal plane while the angle between them remains fixed. The copper vessel P contains apiezon oil; Q is a baffle to deflect splashed oil back into the vessel.

§ 5. Operation

During heating the various parts of the device are as shown in figure 1, the crucible being within the heating element and radiation shields and the whole enclosed in an evacuated water-cooled cover. The temperature is measured optically by sighting through a glass window on the top of the cover and a hole

in the upper radiation shield on to the hole in the crucible. When it is desired to quench, the cam handle is smartly moved by hand through a right angle. Movement of the cam raises the rod H and consequently the crucible to the position shown in figure 2. Simultaneously the top radiation shields are raised at the same speed since they are rigidly connected to the bush G. The side arm M also engages the slot in the cam L and as the arm I moves upwards the arms M and N rotate. The adjustments of the various parts of the device are such that when the crucible just reaches its maximum height it is knocked off the platform C into the quenching tank. As the crucible turns over the loosely fitting stopper and specimen fall out into the oil. The time taken for the complete operation is a fraction of a second. The furnace current is then shut off.

The apparatus has proved to be very reliable and easy to manipulate and no failures of the parts to function have occurred during heat treatment and quenching of more than one hundred specimens. The power consumption is relatively low being approximately 500 watts for 2000°C.

§ 6. Conclusions

An apparatus has been devised for heat treating and oil quenching small specimens (0·5–1·0 g) in vacuum from temperatures up to 2000°C. The temperature of the specimen is measured under approximately full radiator conditions. The furnace has a low power consumption and the apparatus is reliable and simple to manipulate.

Acknowledgments

The apparatus has been constructed as part of a research programme of the National Physical Laboratory on the constitution of titanium alloys, and this description is published by permission of the Director. The author wishes to acknowledge the assistance given by Mr. A. E. Bacon in the use of the apparatus.

References

Alberman, K. B., 1950, *J. Sci. Instrum.*, **27**, 280.
Buckley, H., 1934, *Phil. Mag.*, **17**, 576.
Schofield, T. H., and Bacon, A. E., 1954, *J. Inst. Met.*, **82**, 167.
Schramm, C. H., Gordon, P., and Kaufmann, A. R., 1950, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **188**, 195.

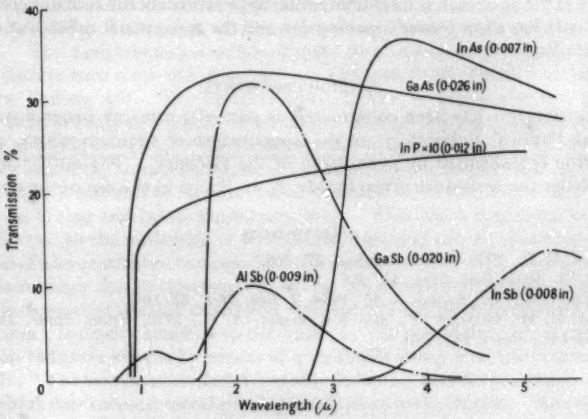
LETTERS TO THE EDITOR

Measurements on some Semiconducting Compounds with the Zinc-Blende Structure

Measurements have been made on the transmission of infra-red radiation through polished specimens of AlSb, GaSb, InSb, InAs, InP and GaAs.

These compounds of elements of Group III and Group V are semiconductors and have the zinc-blende crystal structure. They have all been prepared by fusing the elements† together in an evacuated, sealed-off, silica tube.

Approximate values of the optical activation energy have been deduced from the transmission curves shown in the figure. In the case of AlSb we have not found a well-defined transmission edge. Several samples of this material have shown the same general features but transmission in the region $0{\cdot}7\,\mu$ to $1{\cdot}5\,\mu$ has varied from sample to sample.



Infra-red transmission of various compounds.

It has been shown that the true energy gap for InSb is about 0·18 ev at room temperature and that the position of the transmission edge depends on the number of electrons in the conduction band (Burstein 1954). Because of this it is considered that the present observations should be repeated with purer materials; however, the values obtained, especially for the melting points and the optical activation energies, tend to confirm the general predictions of Welker (1952), see table.

† Johnson Matthey ' Specpure ' except in the case of phosphorus which is a May and Baker product, 99·9% pure.

Hall constant and conductivity measurements were made on GaSb, InSb and GaAs and the density of electrons and electron mobility determined. These figures are included in the table.

Compound	Density of carriers (n cm^{-3})	Mobility of carriers (cm sec^{-1} v^{-1} cm^{-1})	Melting point (°c)	Optical activation energy (ev)
AlSb	—	—	1060	1·55
GaSb	$1{\cdot}2\times10^{17}$	600	725	0·71
InSb	$1{\cdot}2\times10^{16}$	16000	523	0·36
InAs	—	—	950	0·45
InP	—	—	1050	1·20
GaAs	$4{\cdot}6\times10^{17}$	2000	1280	1·25

The variation of the energy gap of GaAs with temperature is -5×10^{-4} ev deg^{-1} over the range 20°c to 190°c. An estimate of the refractive index of this material is 3·4 made by measuring the variation of transmission with the thickness of the specimen and extrapolating to zero thickness. Reflection measurements gave the value 3·3.

The authors wish to thank the Admiralty for permission to publish this note.

Services Electronics Research Laboratory,
Baldock,
Herts.

F. A. Cunnell.
J. T. Edmond.
J. L. Richards.

19th May 1954, and in final form 16th August 1954.

Burstein, E., 1954, *Phys. Rev.*, **93**, 632.
Welker, H., 1952, *Z. Naturf.*, **7**(a), 744.

REVIEWS OF BOOKS

Handbook of Elliptic Integrals for Engineers and Physicists, by P. F. BYRD and M. D. FRIEDMAN. Pp. xii + 355. (Berlin: Springer-Verlag, 1954.) DM. 36 (DM. 39.60 bound in linen).

The authors have collected together for easy reference a very large number of important formulae to facilitate the evaluation of a class of integrals often met with in applied mathematics and physics. They are not concerned with the formal theory of elliptic functions but content themselves with clear definition and the inclusion of enough theory to make the book useful to those who have not met these functions before. In all there are about 3000 formulae and integrals.

The reader is introduced to the three kinds of elliptic integral, Jacobi's elliptic functions and their inverses and to Jacobi's zeta function, all of which are defined and exhibited graphically. A few conformal mappings involving these functions are also given.

A large section is devoted to the classification of integrals involving the square root of a cubic or a quartic or of certain simple polynomials of higher degree. There are also lists of single and multiple integrals of various combinations of Jacobi's elliptic functions; integrals involving trigonometric or hyperbolic functions, which can be evaluated in terms of elliptic functions, are also dealt with. Another section of the book is concerned with integrals and derivatives with respect to the modulus or argument, and with series and product expansions. Weierstrassian elliptic functions and theta functions are considered in an appendix.

At the end of the book tables are given of the complete and incomplete elliptic functions of the first and second kinds, and also of the zeta function.

A. N. GORDON.

Fluorescence Analysis in Ultra-Violet Light, 4th Edn, by J. A. RADLEY and J. GRANT. Pp. xiv + 560. (London: Chapman and Hall, 1954.) 52*s.* 6*d.*

The original edition of this book provided one of the first collective accounts of the use of fluorescence in analysis and identification of materials. Since then utilization of fluorescence phenomena has become very extensive in many different fields and it becomes increasingly difficult to provide a single textbook which can cover the whole subject. In fact the value of the present new edition (the first since 1939) to the expert in a particular field of fluorescence analysis is open to question. In attempting to cover all aspects in an encyclopaedic way the approach to any specific matter is rather limited and often uncritical, though the position is alleviated by provision of copious and reasonably up-to-date references for each chapter.

The book is arranged in two parts. The first section begins with a chapter on electromagnetic radiation, including fluorescence, which to most workers in the subject will appear trivial in content. There follows a chapter on light sources with the main emphasis on high-pressure mercury discharge lamps of proprietary brands. The low-pressure mercury lamp (2537 Å radiation) is given a surprising dismissal in a few words and classed as " of only limited

interest for fluorescence analysis", an incorrect statement, particularly with respect to fluorochromatography. No mention is made of the compact source hydrogen discharge lamp, now available commercially, which gives a continuous emission over much of the ultra-violet region and is of great interest to those with a more exact and spectroscopic approach to fluorescence analysis. Chapter III deals with the various kinds of optical filters for use with light sources etc. In this, as in other chapters, one is struck by the unscientific nature of some statements, e.g. on page 29, with respect to radiation injurious to the eye etc., we read that " the effect of wavelengths longer than this (3050 Å) is successfully opposed by the physiological action of the cells " !

In the next chapter, which reviews the methods of measuring radiation intensity and dosage, most of the text is devoted to photochemical methods rather than to physical methods, which in the more exact studies of ultra-violet radiation and fluorescence have replaced the older methods. Among the few physical methods described is the use of the blackening of zinc sulphide as a dosage indicator. It is apparently not known to the authors that the effect only occurs when water vapour is present.

The first section ends with a description of methods and techniques which is sometimes trivial and includes repetition of earlier material.

The second half of the text presents a series of chapters on the plethora of fields in which fluorescence analysis finds applications, from agriculture and sewage to medicine, food, fuels, minerals, gems and museum research. Here the reader is presented mostly with a series of short statements on specific applications but is supplied with a numerically impressive set of references for each chapter. The text has not always been brought up to date, for example in the case of inorganic materials (boric acid, zinc oxide, zinc sulphide) or in the chapter on minerals (e.g. willemite, ruby and sapphire, diamond). In addition, in the examples cited the text is confused and sometimes self-contradictory. However, this last section covers a very wide range of applications and provides a valuable compendium for initial enquiry into a specific field. At the end of the text a number of photographs are given which show the value of visual observations in fluorescence analysis. Author and subject indices are provided.

As a general observation it can be said that fluorescence analysis has now become so widely applied in a detailed and a quantitative form that a volume on the subject would be more valuable if its several chapters were contributed by experts in the particular fields or, at least, if such informed workers could offer opinions on the subject matter of such a book before its publication.

G. F. J. GARLICK.

Structure and Properties of Solid Surfaces, edited by R. GOMER and C. S. SMITH. Pp. xiii + 491. (Chicago : University Press; London : Cambridge University Press, 1954.) 64*s*.

This book consists of an edited series of papers (together with discussions) which were presented at a Conference organized by the U.S. National Research Council and held in September, 1952. It is excellently produced and, although of small bulk, is a fairly formidable volume of almost 500 closely but clearly printed pages leaving little of the subject untouched. It is naturally more of a review and progress report than a textbook, but the editors have, nevertheless, succeeded in knitting together the subject matter to produce a coherent and well-ordered treatment.

In an introductory, paper Ewald discusses the importance of surface phenomena and stresses the pressing need for an adequate yet simple theoretical model of surface structure. This is followed by an outstanding article by Conyers Herring which is both a review of the classical thermodynamic treatment of surfaces (beginning with Willard Gibbs) and an exposition of the applicability of classical macroscopic concepts to current problems. The important distinction between surface tension and surface stress in the case of solid surfaces is treated admirably. One of the weakest links in the theory is seen to be our rather scanty knowledge of those processes which limit the bulk movement of material under the action of surface forces. This article is of wide scope and provides an extremely sound introduction to the book.

Ewald and Juretschke introduce the atomistic treatment of surfaces, and stress the difficulty of defining a surface for theoretical calculations except on the basis of point charges: the quantum mechanical approach also necessitates fairly drastic simplifications. A critical account of the interpretation of typical experiments on the mechanical properties of metal surfaces under various conditions of equilibrium is given by Shaler. Weyl discusses the influence of ionic polarizabilities on surface properties, but his article is not entirely free from obscure reasoning.

Among the papers of a more experimental nature, G. P. Thomson contributes a short article on electron diffraction methods and discusses the types of information they may be expected to reveal. Thomson does not consider the evidence for anomalous spacings of surface layers to be conclusive; attention is frequently drawn throughout the book to this important question and unequivocal experimental results are much to be desired. Bowden and Tabor's article on adhesion is notable for its account of their exploitation of the reflection electron and shadow-tarnishing microscopic techniques, and a brief report of recent work on the adhesion of molecularly flat mica surfaces deserves attention.

Wells and Buckley contribute papers on topics relating to crystal growth. Much of Wells' article, however, has little direct connection with surface phenomena but will interest the crystal chemist. The arguments which Buckley brings against Frank's theory of spiral growth can scarcely be taken seriously. His unquestioning belief in the physical capabilities of vortices of unexplained origin, which tends towards naivity, is vigorously and justifiably attacked by Cabrera in the discussion. Epitaxy is treated comprehensively by Seifert who points out that detailed knowledge of the forces involved is still lacking.

The remainder of the book is devoted to gas adsorption and catalysis and one gains the impression that further work in this field will be of the greatest value in elucidating many of the outstanding problems of surface action.

Although the authors write with authority, some of them digress unnecessarily or fail to relate their conclusions to the central theme. However, the volume is, on the whole, above serious criticism. The comprehensiveness of its content, together with the frankness and detail of the reported discussions, render it a valuable source of information, particularly with regard to current thought and trends in this field.

The book is very fully referenced and can be strongly recommended.

H. D. K.

Physics and Applications of Secondary Electron Emission, by H. BRUINING. Pp. xii + 178. (London: Pergamon Press, 1954.) 25*s*.

Apart from its fundamental interest in the physics of metals, secondary emission has to-day an enormous number of practical applications, especially in modern electronic equipment. The result is that the literature on the subject is extensive and is rapidly expanding; this applies both to the academic and technological papers. The present book is one of a series of monographs which report on research, and help to bring together the vast amount of data now being published. A modern book on secondary electron emission is necessary and the present work, by Dr. H. Bruining, of the Philips Research Laboratory, Eindhoven, himself an authority, is to be welcomed.

A very wide field is covered in 178 pages. A chapter is devoted to experimental methods of measurement of secondary emission and this is followed by two chapters which review previous results on secondary electron emission from metals and metal compounds. A short chapter is devoted to the variation of secondary emission yield caused by external absorption of ions and atoms, a field of increasing importance; but it is unfortunate that it has not been found possible to include mention of the most recent work in this particular field. Two important chapters are then devoted to theories of electron emission, and the last three deal with applications and these will be particularly valuable to electronic engineers.

The author gives an account of published data and of the theories which have been proposed to account for them. When these are not entirely satisfactory, or when the data themselves are not self-consistent, the author simply states the facts, and this makes the book a valuable work of reference. An extensive list of references is given of important papers up to 1952. The book, should prove extremely useful, especially to those working on electronics, and can be strongly recommended.

F. LLEWELLYN JONES.

The Physics of Experimental Method, by H. J. J. BRADDICK. Pp. xx + 404. (London: Chapman and Hall, 1954.) 35*s*.

It has always been true of the experimental sciences, and particularly of physics, that advances in the knowledge of fundamental processes have been closely linked with technical improvements in the basic apparatus available to the experimenter. This is well shown by the outstanding progress in the present century, which has been undoubtedly due in no small measure to the gradual development of vacuum technology and high voltage electrical engineering, to the arrival on the scene of the radio valve and the corresponding emphasis on high-frequency circuit techniques, and to the availability of a wide range of instruments for the detection of radiations of all kinds.

It is unfortunately rather uncommon in University courses for the student to meet with more than incidental references to such technical matters. Such references as do occur are chosen partly by custom, partly because of the closeness of the subjects to pure physics, and partly by the criterion of how conveniently they can be used by a lecturer as illustrations of physical principles. From the postgraduate point of view the situation is often just as unsatisfactory. With three short years in hand and an original contribution to physics expected, a Ph.D. student is usually more interested in starting his experiments than in

indulging in a critical survey of the properties and limitations of the materials, techniques and instruments available to him, however beneficial such a survey might be in the long run. Such information is normally picked up in a rough and ready fashion during the course of his work, more often than not by making mistakes and so wasting time.

Dr. Braddick in his book brings together successfully much of this basic practical knowledge, and by discussing at the same time the fundamental principles involved makes it easy to appreciate and remember. The range covered is, as would be expected, quite wide. The author begins by giving an account of the mathematical treatment of results, statistical methods, errors, curve fitting etc., with a useful appendix on numerical methods. Following this are chapters on the mechanical design of apparatus and on the properties of the materials normally used, such as metals, alloys, solders, magnetic materials, plastics, glasses etc., containing much useful data. Next come vacuum techniques, electrical measurements, electronics (covering the normal circuits used in laboratory practice), optical systems, and photography. A very interesting chapter follows in which fluctuation phenomena are discussed in connection with the inherent limits of measurement, illustrated by noise in electrical circuits, fluctuations in galvanometer readings, and the detection of weak light beams by the photocell and electrometer triode. The last chapter is concerned with the special techniques used in nuclear physics, including ionization chambers, Geiger counters, scintillation counters, the cloud chamber, and the photographic emulsion as a means of particle detection.

The treatment throughout is directed at giving an experimenter who has a problem to solve some guidance in the choice of instruments, the design of apparatus, and the assessment of errors and limitations. In this way it becomes less a reference book and more a handbook, a collection of information and principles which should be (and eventually will be) known by any successful research worker in physics, and indeed in many allied fields.

As to the author's choice of subject matter, it would be fair to say that most of the topics will be of use at some time or another to a physicist. The only noticeable omissions concern the methods of thin film production, i.e. windows, thin targets and surface films, and the practical details of sources of emission such as cathodes, ion sources etc. In these particular cases the availability of either the window or current of particles is often a vital factor in the feasibility of a proposed experiment.

The book is remarkably free from errors considering the amount of information it contains, and is well supplied with references for further reading.

R. LATHAM.

Fields and Waves in Modern Radio, by SIMON RAMO and JOHN R. WHINNERY. 2nd Edn. Pp. viii + 576. (London: Chapman and Hall, 1953.) 70*s*.

The new edition of *Fields and Waves in Modern Radio* differs from the first in many respects. Rationalized M.K.S. units have now been used throughout to bring the book in line with modern practice, there has been a general revision of all material to improve and clarify presentation, and many new aspects have been introduced which were not included in the first edition. The final result is very pleasing.

The authors have had considerable experience in teaching microwave theory and have themselves contributed to its development. Their style of presentation is leisurely and readable and in keeping with their desire to provide a comprehensive text for intermediate students. The first four chapters are devoted to the development of the fundamental laws of electromagnetism. Each new aspect is approached cautiously. Great importance is attached to the development of physical pictures of the simpler phenomena which will provide a sound basis for the more difficult problems to come. The first chapter is concerned with simple L–C circuits and transmission lines; the purpose is, however, to introduce in their simplest form oscillating systems, wave propagation and several mathematical techniques.

In the next chapter the student is introduced to field concepts. Again the policy is to discuss and undertake simple aspects first, and with this in view the discussion is confined to stationary fields. The authors realize the danger in proceeding from the particular to the general case, but issue repeated warnings when discussing laws which apply only to stationary fields. This chapter also introduces vector notation and Laplace's equation, to be followed in the next chapter by a discussion of mathematical techniques for the solution of the latter. In Chapter 4 Maxwell's equations for time-varying fields are introduced for the first time.

The remainder of the book is devoted to the application of the field equations to electromagnetic problems. Since in many practical cases exact solutions cannot be found, it is important that methods of approximation should be well understood. In microwave problems the concept of lumped circuits is often used and this technique is discussed at some length. The text then moves on to a number of field problems which have exact field solutions, such as the propagation and reflection of electromagnetic waves, and the elementary theory of waveguides and resonant cavities. This is followed by a chapter on coupling from one system to another and its representation by microwave networks. In the final chapter the radiation field of microwave antennae is discussed; these include dipoles, linear arrays, slot radiation and the radiation from various apertures.

W. WALKINSHAW.

Acoustics; a Text-Book for Physics and Engineering Students, by T. M. YARWOOD. Pp. x+346. (London: Macmillan, 1953.) 15*s*.

Of the considerable number of elementary books on sound which have appeared in post-war years, this one merits honourable mention. Written by one who has considerable experience in teaching physics, it is well adapted to the needs of the higher forms of grammar schools and intermediate classes in Universities and technical colleges. The book is ably illustrated with plates and line drawings (though it is true that many of the latter seem vaguely familiar to a constant reader in acoustics and make him wonder whether publishers have a pool for such figures!). The mathematical equipment required of the student is modest and he is assisted by a good supply of examination problems. The author has evidently been at some pains to make the book up-to-date and to stimulate interest in those parts of the subject which lie outside the usual examination syllabus.

E. G. R.

CONTENTS OF SECTION A

PAGE

Dr. S. Hochberg, Prof. H. S. W. Massey and Mr. L. H. Underhill. The Scattering of Nucleons by Alpha Particles—the s-Phases 957

Dr. B. Davison and Mrs. M. E. Mandl. On the Neutron Spectrum for $\nu^{-\alpha}$ Absorption Cross Section 967

Dr. F. C. Flack, Dr. J. G. Rutherglen and Dr. P. J. Grant. Radiations from the Proton Bombardment of ^{23}Na 973

Mr. I. P. Grant. Theory of (d, p) and (d, n) Reactions: I—General Theory Ignoring Coulomb Effects 981

Dr. W. M. Fairbairn. The D–D Reaction 990

Dr. F. J. M. Farley and Mr. J. R. Storey. The Sidereal Correlation of Extensive Air Showers 996

Dr. G. N. Fowler. Note on the Relation between certain Forbidden Beta Transitions and the Nuclear Spin Orbit Interaction 1005

Dr. A. Dalgarno. Inelastic Heavy-Particle Collisions Involving the Crossing of Potential Energy Curves: II—Charge Transfer from H-atoms to Al^{3+}, B^{2+}, Li^{2+} and Al^{2+} 1010

Research Notes:

Dr. R. H. Tredgold. On Antiferromagnetism in Metals . . . 1018

Mr. L. S. Kothari. Riesz Potential and the Elimination of Divergences from Quantum Electrodynamics: III 1021

Dr. G. D. Adam and Dr. K. J. Standley. Ferromagnetic Properties of Oxidized Mn_2Sb 1022

Letters to the Editor:

Mr. G. R. Satchler. Radiative Transitions in j–j Coupling . . . 1024

Mr. J. G. Dardis and Prof. C. B. A. McCusker. A Possible Variation of the Rate of Local Penetrating Showers 1026

Reviews of Books 1028

Contents of Section B 1036

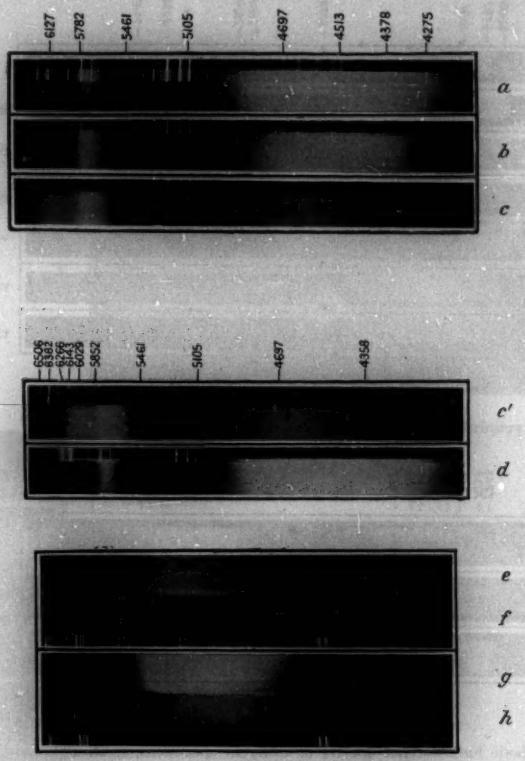


PLATE I.

a–d. Cathode luminescence spectra of different specimens of MgO crystal taken with a glass spectrograph. Exposure time 25 min each for *a*, *b* and *c*, 2½ hours for *c'* and 30 min for *d*.

Nature of the specimen.

a. Fresh; *b*, *c*, and *c'* had previous exposure of 1 hour *b*, 10 hours *c*, and 30 hours *c'* to cathode rays; *d*, same as for *c'* but heated in a Bunsen flame after exposure to cathode rays.

e–h. Comparison of phosphorescence and fluorescence spectra of two specimens of MgO crystal taken with a baby quartz spectrograph using a wide slit and a Hartmann diaphragm; *e*, phosphorescence spectrum of a fresh crystal; exposure time 25 hours; *f*, fluorescence spectrum of the above; exposure time 2 sec; *g*, phosphorescence spectrum of the same specimen as used for *d*; exposure time 25 hours; *h*, fluorescence spectrum of the above specimen as used for *g*; exposure time 2 sec.

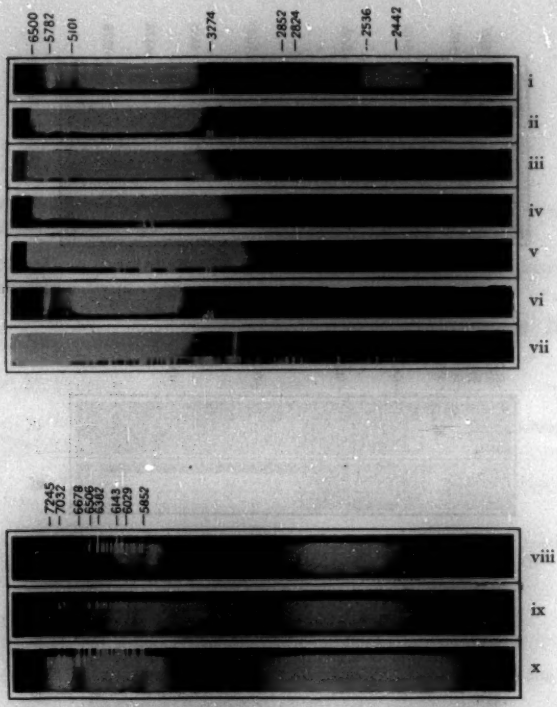


PLATE II.

i–vi. Cathode luminescence spectra of different specimens, fresh and heated, of MgO crystal taken with quartz spectrograph.

i. Fresh crystal heated in oxy-coal-gas flame for 3 minutes and then cooled; exposure time 10 min.

ii. Specimen as used in i but having a previous exposure to cathode rays of 20 hours after being heated in the oxygen coal-gas-flame; exposure time 10 min.

iii. Fresh crystal heated in Bunsen flame for 5 hours and then cooled; exposure time 10 min.

iv. Crystal previously exposed for 20 hours, heated in a Bunsen flame for 30 min, and again exposed to cathode rays for 10 hours; exposure time 10 min.

v. Crystal previously exposed to cathode rays for 20 hours, exposure time 10 min.

vi. Fresh crystal; exposure time 25 min.

vii. Emission spectrum of the MgO crystal when heated in an oxy-coal-gas flame exposure time 40 sec.

viii, ix, x. Cathode luminescence spectra taken with glass spectrograph using plates sensitized in the region 7000–8000 Å. Crystal specimen for spectrogram viii same as for iii; ix same as for v; x same as for ii.

(i) NaI (ThI) (ii) $CaWO_4$

(*a*) Typical photographs of secondary pulses. Primary is clearly visible in (ii).

$T = -45°C$

$T = -31°C$

(*b*) Primary pulse envelope at three temperatures with $CaWO_4$, showing increase in peak height.

$T = -18°C$

Figure 2.

Figure 9. Magnification ×750. Etched with ammoniacal 10% ammonium persulphate solution with hydrogen peroxide. Widmanstätten structure in heat-treated alloy is shown.

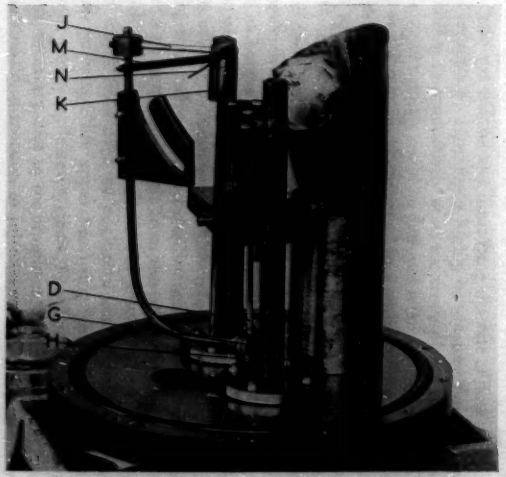


Figure 1. Apparatus for heat treatment and quenching. Heating position. (×7/20 approximately.)

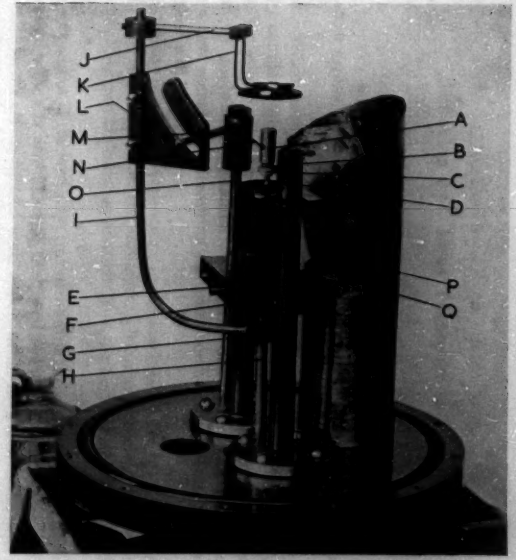


Figure 2. Apparatus for heat treatment and quenching. Quenching position. (×7/20 approximately.)

Diffusion and Mass Transport in Tubes

By G. I. TAYLOR
Trinity College, Cambridge

38*th Guthrie Lecture, delivered* 24*th September* 1954; *M.S. received* 29*th September* 1954

Abstract. When soluble matter is introduced into a solvent flowing slowly through a capillary tube it is dispersed longitudinally by a process which involves both the variation in fluid velocity over the cross section of the tube and radial diffusion by molecular agitation. Measurements of longitudinal dispersion provide a new means for measuring diffusion coefficients. Results obtained by this method will be given.

The stability of solutions contained in vertical tubes when the density increases upwards also depends on radial diffusion. Measurements of the equilibrium density gradient can be used as another new method for determining diffusion coefficients.

The mechanics of dispersion in turbulent flow through a pipe can be discussed by a method which is analogous to that used for streamline flow. The results of this calculation are compared with experiments in which brine was injected into water flowing in 3/8 inch and 40 inch pipes, and the subsequent dispersion along the pipe was measured. Similar comparisons are made with American measurements in long pipe lines.

About two years ago Dr. Mount of the Animal Physiology Laboratory at Babraham told me that he was measuring the rate of flow of blood in the arteries of animals by injecting a highly conducting fluid at a point and observing the variation in conductivity at an electrode placed downstream. He asked me how the mean velocity could be deduced from his observations. If the injected material would remain concentrated in a small volume at the middle of the artery it would move with the maximum velocity u_0 of the blood stream and after a time x/u_0 the conductivity at an electrode distant x downstream would suddenly rise. By measuring this time the value of u_0 would be obtained. In fact the injected material does not remain concentrated; it spreads out along the artery and the conductivity downstream rises gradually, reaches a maximum and then decreases to its normal value.

The problem which confronts anyone who wants to use this method for measuring the rate of flow, or mean speed, is, therefore, to determine which point on the conductivity–time curve, obtained at a distance x downstream, corresponds with the arrival of a point which travels from the point of injection to the point of observation with the mean speed of flow. In practice everyone who has used the method, and it has been used by engineers and physicists as well as by physiologists, has calibrated his apparatus using an independent means of measuring flow rate. In complicated systems, like those dealt with by physiologists, calibration is necessary, but in the simple case of flow through a straight pipe the mechanics of the dispersion of injected material can be analysed and understood.

Some of these calculations are now published and some will soon appear in the Royal Society's *Proceedings*. Here I will attempt to explain their physical basis and compare them with the result of experiment rather than go into mathematical analysis. The cause of dispersion in a pipe is the variation in velocity over its cross section. When the flow is not turbulent the dispersion produced by convection is modified by molecular diffusion. When it is turbulent the convective dispersion is modified by turbulent diffusion.

The first step in understanding these processes is to consider dispersion by convection alone. Imagine a straight pipe through which a viscous fluid is streaming in non-turbulent flow and suppose that at time $t=0$ the colour of the stream entering at $x=0$ is changed, say by letting in a dye instead of a pure water.

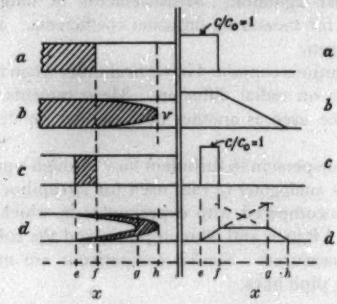


Figure 1. Figure 2.

Figure 1. Dispersion by convection alone. (*a*) and (*c*) are initial distributions, (*b*) and (*d*) distributions at time t.

Figure 2. Distribution along tube of mean concentration over cross sections corresponding with cases shown in figure 1.

The values of x in figures 1 and 2 are $-L$ at e, 0 at f, $2Ut-L$ at g and $2Ut$ at h.

The initial condition is indicated in figure 1 (*a*). The distribution of velocity is parabolic and the mean velocity U is half the maximum, so that after time t the forward edge of the colour lies in a paraboloid whose vertex is at $x=2Ut$. This condition is shown in figure 1 (*b*). Now suppose our measuring instrument measures the mean concentration, c, of the dye at any section. The areas of sections of a paraboloid are proportional to their distance from its vertex. If c_0 is the concentration of the dissolved dye as it enters the tube the mean concentration at any section is

$$c=c_0\left(\frac{\text{area of section of paraboloid}}{\text{area of cross section of tube}}\right).$$

Thus the concentration decreases linearly from $c=c_0$ at $x=0$ to $c=0$ at $x=2Ut$. The initial distribution of c is shown in figure 2 (*a*), and the distribution at time t in figure 2 (*b*). Next consider the case when the dye is initially confined to a short length of the tube from $x=-L$ to $x=0$, the fluid in front and behind being pure solvent. The condition is represented in figure 1 (*c*). After the time t both the front and back surfaces of separation are paraboloids (figure 1 (*d*)). The dye has penetrated to $x=2Ut$ and from $x=2Ut-L$ to $x=2Ut$. c decreases exactly as in figure 2 (*b*). c is constant from $x=0$ to $x=2Ut-L$, and from $x=-L$ to $x=0$ increases uniformly. The distributions of c at times 0 and t are shown in figures 2 (*b*) and 2 (*d*).

To obtain these distributions of concentration experimentally it is necessary to carry out the experiment so quickly that transverse diffusion does not have time to modify the dispersion produced by pure convection. We found no difficulty in doing this, using apparatus constructed for use also at slow speeds of operation. This apparatus is shown diagrammatically in figure 3. Here A and B

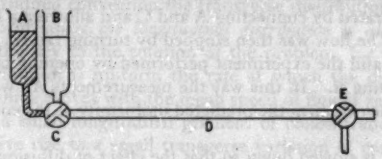


Figure 3. Apparatus for observing dispersion in a tube.

are two receptacles. A is filled with a soluble substance whose dispersion is to be measured. B is filled with the solute, in our case distilled water. A and B are connected with the capillary tube D through a 3-way glass tap C. To measure the concentration of a dissolved substance in a capillary tube is not always easy. In our experiments a strongly coloured salt, such as potassium permanganate, was used and the distribution of concentration was measured by cutting off a piece of the capillary and filling it successively with solutions of known concentration. The points on the main tube at which the colour matched that of the comparison tube were observed. In this way the concentration c was measured as a function of x, the distance along the tube from the 3-way tap. To operate the apparatus A was first connected with D so that the full concentration c_0 of the solution in A was in contact with the tap C. Tap C was then turned so as to flush out the tube D with water from B. To perform the experiment tap C was again turned so as to connect A with D. After a measured interval of time

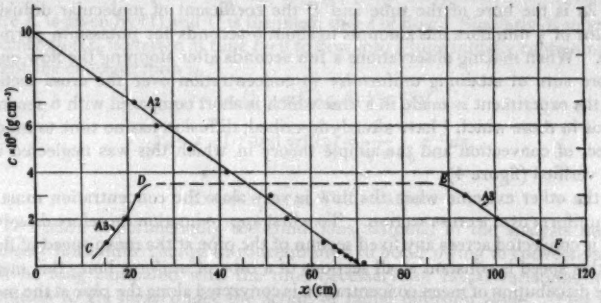


Figure 4. Measurements of c when dispersion occurred in $1\frac{1}{2}$ seconds.

the flow was stopped, usually by closing a tap E at the far end of D. It was found that the longitudinal molecular diffusion is so slow that the distribution of concentration did not change by a measurable amount in several hours so that there was plenty of time to make the measurements.

Figure 4 shows the results of two experiments in which total duration of the flow was only $1\frac{1}{2}$ seconds. The points numbered A2 represent measurements

in which the apparatus was operated in the manner already described. I have already shown that according to the simple theory which considers only convection the distribution of concentration should in this case decrease linearly, and A2 in figure 4 shows that this is in fact the case, a straight line passes very close to the observed points.

To observe the dispersion of an initially concentrated volume of solute the apparatus was operated by connecting A and C and allowing a little of the solute to flow into D. The flow was then stopped by turning tap E. Tap C was then connected with B and the experiment performed by opening tap E for a known time and then closing it. In this way the measurements A3 were obtained. It will be seen that they confirm the predicted uniformity of the concentration along most of its extent.

When the flow is slowed down so that the effect of diffusion is not negligible the simple convection theory breaks down and the solute is not dispersed along the tube nearly so quickly. The reason for this can be understood by considering figure 1 (*b*). The vertex of the paraboloid at v is carried by convection into the pure solvent. The solute then diffuses radially from it so that the front of the patch of solute is continually getting less concentrated than the simple convection theory would predict. On the other hand the portion near the wall at the rear end of the patch is continually being left behind. The solute therefore diffuses into the middle of the tube and is then convected forward into the patch of solute again. It is difficult to analyse this process mathematically except in the case when the solute has spread so far along the tube that it occupies a length which is many times its bore. For this reason I have limited my studies to that case. Without going into the details of this calculation, which have already been published (Taylor 1953, 1954 a), I can explain the physical ideas behind it. First one notices that if the bore of the experiment tube is small the time taken for radial diffusion to equalize any radial variation in concentration is also small, in fact the most persistent radial variation dies to 1/e or 0·37 of its value in time $a^2/14{\cdot}4D$, where $2a$ is the bore of the tube and D the coefficient of molecular diffusion. In a tube of $\frac{1}{2}$ mm bore this amounts to about 6 seconds for potassium permanganate. When making observations a few seconds after stopping the flow one is therefore sure of attaining uniformity in concentration over the cross section. When the experiment is made in a time which is short compared with 6 seconds, as it was in those which I have already described, diffusion has no time to modify the effect of convection and the simple theory in which this was neglected was in fact verified (figure 4).

At the other extreme when the flow is very slow the concentration must be nearly uniform over a cross section. To a first approximation therefore dissolved matter is convected across any fixed section of the pipe at the mean speed of flow. Since this speed is constant at all sections of a tube of uniform bore, this means that the distribution of mean concentration is convected along the pipe at the mean speed of flow without change of form. To me this seemed a remarkable conclusion because the solvent in the centre of the tube moves twice as fast as the mean speed, so that pure solvent situated on the centre line must catch up the solute and first absorbs and then rejects the dissolved substance, to pass on as uncontaminated as it was before entering the contaminated zone. On mentioning this, to me surprising, result to Professor George Temple, he exhibited no surprise at all and told me that when he was working as a research student at Birkbeck

College with Dr. A. Griffiths on the viscosity of water at very slow rates of flow the mean velocities were measured by injecting a spot of dye into the water as an index. This index was found to remain coherent and was assumed to move at the mean speed of flow. Dr. Griffiths' argument (Griffiths 1911) on this point is a little obscure, but I think it runs on the lines given above.

When the transverse diffusion is not sufficiently great to wipe out altogether the effect of longitudinal convection the transverse distribution of concentration will not be uniform *unless the longitudinal distribution is also uniform.* This remark contains the key to the solution of the problem, for we see that if the longitudinal distribution is uniform the rate at which the dissolved substance passes a section which moves with the mean speed of flow is zero. On the other hand, if there is a small longitudinal gradient of concentration along the tube convection will give rise to a small transverse variation of concentration which in turn will give rise to a small transport of the solute across a section which moves with the mean speed. It is not a long step then to see that this small transport and the small longitudinal concentration gradient must be proportional to one another. Thus the combined effect of longitudinal convection and transverse diffusion is to disperse the solute longitudinally relative to a frame moving at the mean speed of flow by a mechanism which obeys the same law as ordinary one-dimensional diffusion relative to a fluid at rest. The virtual coefficient of diffusion, K, for this process is calculated to be (Taylor 1953)

$$K = \frac{a^2 U^2}{48D}. \qquad \ldots\ldots(1)$$

Thus all the calculations which have been made in the past about diffusion in, say, a diffusion cell can be applied directly to longitudinal dispersion in fluid flowing through a tube. In particular we can predict that if a mass M of a soluble material of constant diffusibility K is initially concentrated at $x=0$ in a small length of tube of radius a the concentration after time t at distance x will be

$$c = \tfrac{1}{2} M a^{-2} \pi^{-3/2} K^{-1/2} t^{-1/2} \exp\left[-(x-Ut)^2/4Kt\right] \qquad \ldots\ldots(2)$$

where K is given by (1) and U is the mean speed of flow. Similarly, if solution of concentration c_0 is started at time $t=0$ to flow into a tube initially containing pure solvent the concentration is

$$c = \tfrac{1}{2} c_0 \{1 - \operatorname{erf}(\tfrac{1}{2} x_1 K^{-1/2} t^{-1/2})\} \qquad \ldots\ldots(3)$$

where $x_1 = x - Ut$ and $\operatorname{erf}(z) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^z e^{-z^2} dz.$

Experimental Verification

To realize experimentally the conditions assumed in the analysis the time necessary for a finite change in concentration to occur owing to convection alone must be long compared with the time necessary for radial variations in concentration to die away under the action of radial diffusion alone. This means that with a tube $\frac{1}{2}$ mm diameter the speed of flow must be of order 1 to 10 cm per minute. At such low speeds it is necessary to devise methods for controlling the flow at a constant rate.

A constant small pressure difference between the end of a tube is difficult to obtain, and in cases where the solution has a different viscosity from the solvent, constancy of pressure difference between the ends of a tube would not ensure

constancy of rate of flow. In later experiments constancy of flow was obtained by fixing apparatus which removed fluid from the downstream end of the tube at a constant rate. One obvious way to do this is to fix to the downstream end of the experiment tube a capillary of much smaller bore. Another is to connect it with a cylinder whose volume is altered at a constant rate by a piston operated by a micrometer screw. Figure 5 shows the measured distribution of potassium

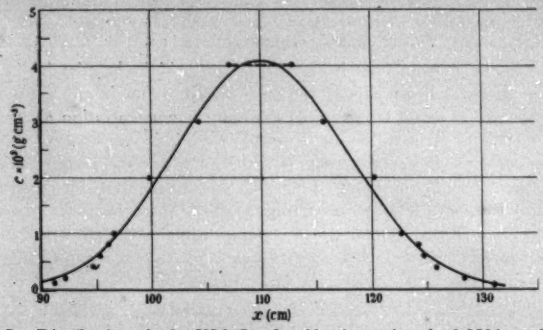


Figure 5. Distribution of c for $KMnO_4$ after 11 minutes in tube 0·0504 cm diameter.

permanganate in a tube 0·0504 cm diameter after the flow had been going for 11 minutes. Initially the colour was concentrated in the first few centimetres of the tube. It will be seen that the distribution is very close to the error curve which is marked. The dispersion is therefore as predicted and, comparing the parameter of the error curve in figure 5 with the formula (2), it was found that $K = 0{\cdot}0459$ cm^2 sec^{-1}. Since $U = \frac{110}{600} = 0{\cdot}167$ cm sec^{-1}, the diffusion coefficient D, found from (1), is

$$D = \frac{(0{\cdot}167)^2\,(0{\cdot}0252)^2}{48\,(0{\cdot}0459)} = 0{\cdot}80 \times 10^{-5}\ \text{cm}^2\ \text{sec}^{-1}. \qquad \ldots\ldots(4)$$

Though this value lies within the range of previous measurements of D for potassium permanganate, these vary through a large range, in fact for nearly all strongly coloured solutions D is very dependent on concentration. In figure 6

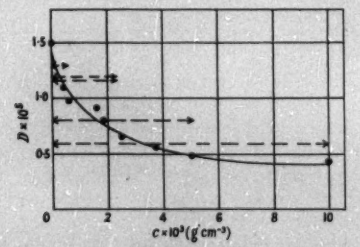


Figure 6. Value of $D = a^2U^2/48K$. Range of values of c in measurements of K shown by broken lines. Fürth and Ullmann's measurements using diffusion cells shown by circles.

the range of concentrations covered by each experiment and the value obtained by applying equations (2) or (3) together with (1) are shown. The experimental results of Fürth and Ullman (1927), who used diffusion cells, are also shown.

It is clear from the comparison shown in figure 6 that the measurement of dispersion might provide a much simpler technique for measuring c than the use of diffusion cells with all their attendant practical difficulties, but when D is not independent of c the formulae (2) and (3) are not accurate. In such cases the only practical methods for measuring c are, either to measure diffusion between pairs of solutions of slightly different concentrations, a method which cannot be used when concentrations are estimated colorimetrically, or to use only the type of experiment envisaged in (3), i.e. to introduce the solution at one end of the tube, which is initially filled with pure solvent, and maintain a constant flow of the solution.

The reason why this type of experiment can be analysed, and experiments in which the diffusion of an initially concentrated mass is measured cannot be interpreted, was originally pointed out by Boltzmann. He showed that when diffusion depends on concentration there is one case in which successive distributions of concentration are all similar to one another, namely that in which the solution and solvent are each initially uniform and are separated by a plane surface. Boltzmann showed that at time t after diffusion started the concentration is a function of $\zeta = x t^{-1/2}$ only, and that

$$D = -\tfrac{1}{2}\frac{d\zeta}{dc}\int_0^c \zeta\, dc. \qquad \ldots\ldots(5)$$

Thus D can be determined for the whole range of concentrations which occur in the experiments if c can be measured accurately as a function of ζ.

I tried applying this formula to my measurements of dispersion to obtain K instead of D as a function of ζ, but when I calculated D from the values of K so obtained I found considerable differences between my values and those given in Landolt and Börnstein's tables. Such is my respect for these famous tables that I thought I must be wrong, but on looking up the original authors quoted I found that without exception all the measurements of diffusion of dyes and strongly coloured substances—covering many pages of the tables—are based on the equation (3), which is only correct when D is independent of c.† To get an idea of how big the error might be I calculated the distribution of concentration for an ideal case in which D varies as $e^{-\beta c}$. I found that when β is chosen so that this ideal variation is near the variation given in Landolt and Börnstein over the range used in the experiment, the error which the authors quoted in those tables could have made was in some cases greater than 50%.

It seems, therefore, that it is worth while to develop the dispersion apparatus so that it can make measurements of concentration accurately enough for use in deducing K from Boltzmann's integral and then finding D from (1). I am now engaged on this work.

Dispersion in Turbulent Flow

The success of the method I have described in predicting the dispersion of a dissolved salt along a pipe through which a non-turbulent stream is flowing encouraged me to try whether the same physical principles could not be applied to turbulent flow. The part played by radial diffusion in the former case must

† This criticism does not apply to all the diffusion measurements reported in Landolt and Börnstein's tables. In other cases Boltzmann's method was used.

be replaced by turbulent radial transport in the latter. Few measurements of radial transport of matter in pipe flow seem to exist, but every measurement of pipe friction is automatically a measurement of radial transport of momentum, and there are vast numbers of these. Though it is not possible to prove an exact connection between the transport of matter and the transport of momentum in turbulent flow, Osborne Reynolds pointed out the consequences which would arise if they were exactly analogous. Reynolds' analogy states that the virtual coefficient of diffusion for momentum is identical with that for heat or for any other transferable property of the fluid. In the case of radial transfer in a pipe this is expressed by the equation

$$\epsilon = \frac{\tau}{\rho\,(\partial u/\partial r)} = \frac{m}{\partial c/\partial r}, \qquad \ldots\ldots(6)$$

where ϵ is called the coefficient of turbulent transport, τ is the turbulent stress at radius r, u is the velocity and m the rate of transfer of matter in a radial direction. If τ_0 is the surface friction on the pipe the equation connecting τ and τ_0 is

$$\tau = \tau_0 r/a. \qquad \ldots\ldots(7)$$

Since τ_0 is measured in pipe friction experiments, equations (6) and (7) provide a connection between m and $\partial c/\partial r$ in terms of quantities previously measured. Though Reynolds' analogy is not demonstrably true, and in fact is demonstrably untrue in the case of free turbulence, it is found to be very nearly true in the case of turbulence near solid boundaries and in pipes, so that there is some justification for assuming it to be true in the present case.

The equation of continuity, that is the equation which expresses the fact that none of the dispersing material is lost during the process of dispersion, can be used in a way which is analogous to its use in the case of molecular diffusion. Considering as before only cases where the dispersing material is spread over a length which is great compared with the diameter of the pipe, it is found as before that, relative to axes which move with the mean speed of flow, material is dispersed along the pipe as though by a virtual coefficient of diffusion

$$K = 10{\cdot}1 a v_*. \qquad \ldots\ldots(8)$$

Here v_* is the ' friction velocity ', which is defined by the equation

$$\tau_0 = \rho v_*^2. \qquad \ldots\ldots(9)$$

A more familiar method for expressing the results of pipe friction experiments is

$$\tau_0 = \gamma(\tfrac{1}{2}\rho U^2) \qquad \ldots\ldots(10)$$

where γ is a non-dimensional friction coefficient, U the mean speed of flow and ρ the density of the fluid. Comparing (9) with (10)

$$v_*/U = \sqrt{(\gamma/2)}, \qquad \ldots\ldots(11)$$

so that (8) can be expressed in the form

$$K = 10{\cdot}1 a U(v_*/U) = 7{\cdot}14 a U\sqrt{\gamma}. \qquad \ldots\ldots(12)$$

This result (Taylor 1954 b) looks very simple, but the ideas behind it and the mechanism it envisages are not so simple. However, the problem of the dispersion of material injected into pipes is one of practical importance, and several experimental investigations have been made which can be compared with the theoretical formula. The earliest was that of Allen and Taylor (1923), who wanted to measure the speed of flow in large pipes conveying water to power plants. They injected salt into a stream flowing through a straight 40-inch pipe

355 feet long and measured the conductivity at the far end. Their object was not to measure the dispersion but to find out what point on the conductivity–time curve corresponds with the time taken by points travelling with the mean speed to cover 355 feet. They measured the mean flow also by other methods and found that the required point on the conductivity–time curve was that at which the concentration was a maximum, as the present theory predicts. They did not discuss the dispersion, but their published conductivity–time curves look like error curves. This is in accordance with the theory which predicts that the distribution of concentration about a point which moves with the mean speed is

$$c = (\text{const})t^{-1/2}\exp(-x_1^2/4Kt) \qquad \ldots\ldots(13)$$

where x_1 is the distance from the centre of the concentration. By fitting the best error curve to one of Allen and Taylor's observed distributions a value $K = 2{\cdot}18 \times 10^3$ cm^2 sec^{-1} was found. In this experiment $U = 105$ cm sec^{-1} and $a = 20 \times 2{\cdot}540 = 50{\cdot}8$ cm, so that, taking the viscosity of water as 0·011, the Reynolds number of the flow was $R = 105 \times 50{\cdot}8/0{\cdot}011 = 9{\cdot}7 \times 10^5$. At this Reynolds number the resistance coefficient is such that $U/v_x = 26$. To compare theory with observation it is convenient to calculate the measured value of $(K/aU)(U/v_x)$, which according to (12) should be 10·1. In the experiment just described this is

$$\frac{K}{av_x} = \frac{2{\cdot}18 \times 10^3}{(50{\cdot}4)(105)}(26) = 10{\cdot}6. \qquad \ldots\ldots(14)$$

The agreement between this and the theoretical value 10·1 is better than one would be justified in expecting from the nature of the experiment, and in fact another of Allen and Taylor's experiments yielded $K/av_x = 11{\cdot}7$ when subjected to the same analysis.

In Allen and Taylor's experiments the measurements from which dispersion can be derived were only incidental to their main purpose. More recently dispersion in pipes has become important because very long pipes are used to convey fluids to great distances. Different grades of oil, say gasoline and Diesel oil, are transported successively in the same pipe. The pipe is not emptied between successive periods of use, so that there is a zone of mixture in which the surface of separation between two miscible fluids is dispersed along the pipe. It is of economic importance to know how much of each fluid is contaminated by mixture with the other. Hull and Kent (1952) have investigated this question by injecting a radioactive substance into a 10-inch pipe 182 miles long and measuring its dispersion at various points along its course when $U = 81{\cdot}7$ cm sec^{-1}. Figure 7 shows the concentration–time curves at stations 13·8, 43·1 and 108·5 miles from the point of injection. Error curves have been superimposed on the observed points, and these have been shifted so that their axes lie on the centre line of the figure. The ordinates therefore represent $t - x/U$. It will be seen that the dispersion is very closely gaussian, but when K is deduced from them it is found that K/av_x is larger than 10·1. It varies from 12·4 for the station nearest to the point of injection to 22·5 at 182 miles. The theoretical formula applies to a straight pipe, so that it is hardly to be expected that it would apply accurately to a pipe which accommodates itself to the hills and valleys over which it passes, but the discrepancy is more than I should have expected in view of the good agreement which was found in other large-scale experiments.

Smith and Schulze (1948) describe measurements with pipes 440 miles long in which two products A and B followed one another. At various points along

the pipe samples were drawn off, and from these the length, S, of pipe in which the mixture changes from 1% A and 99% B to 99% A and 1% B as found. Using the theoretical formula (3) with (12) to calculate S I found that

$$S^2 = 437ax(v_x/U) \qquad \ldots\ldots(15)$$

where x is the length of the pipe between the entry and the point of measurement. The comparison displayed in table 1 between the values of S measured in a 12-inch pipe and those calculated using (15) shows very good agreement.

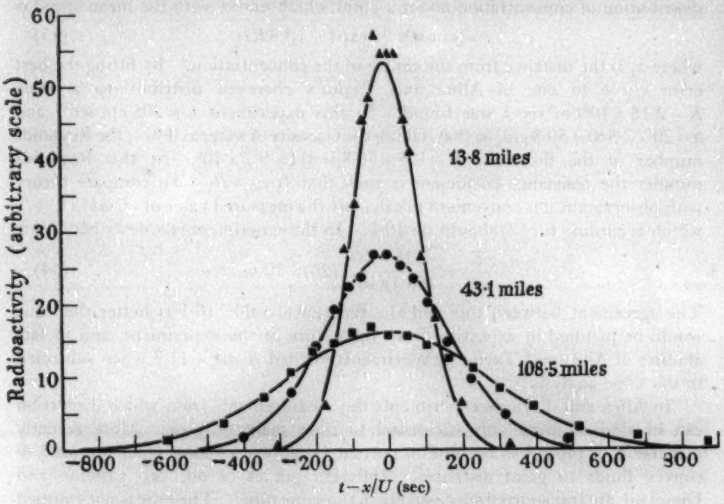


Figure 7. Dispersion in a 10-inch pipe (Hull and Kent 1952).

When I made the theoretical analysis I did not know of the existence of these large-scale experiments, so, with the help of Dr. T. H. Ellison, I set up apparatus in the Cavendish Laboratory to measure dispersion in a pipe 0·9 cm diameter

Table 1. Comparison between Calculated Values of S and those observed by Smith and Schulze (1948) in a Long Pipe Line

x (feet)	$10^{-4}R$	U/v_x	S (feet) equation (15)	S (feet) observed	x (miles)
322186	5·9	25·4	1670	1770	61
661109	5·76	25·4	2390	2425	125
1018301	5·84	25·4	2970	2890	193
1402896	5·91	25·4	3480	3505	265
1726085	5·79	25·4	3870	3895	327
2033803	5·79	25·4	4190	4360	385
2279538	4·50	24·6	4340	4670	432

and 16 metres long. Salt was injected by a spring gun at one point and the conductivity at a point 16 metres downstream was recorded on a rotating drum camera. The conductivity–time curve so found was nearly gaussian in the middle but was not quite symmetrical particularly when the Reynolds number

of the flow was low. This, I think, was due to the fact that in that case an appreciable amount of the brine was in the laminar boundary layer close to the wall where the Reynolds analogy does not hold.

The results of these experiments are given in table 2. It will be seen that at Reynolds number $1{\cdot}2\times10^4$, where the conductivity–time curve deviates considerably from the gaussian form, K/av_x is 12·8, which is well above the theoretical value, but at $R=1{\cdot}9\times10^4$ K/av_x is 10·0, which is very close to the calculated value.

Table 2

Pipe	U cm/sec	R	U/v_x	K/av_x
40-inch (Allen and Taylor)	105	$9{\cdot}7\times10^5$	26·0	10·6
$\frac{3}{8}$-inch smooth $x=322$ cm	222	$1{\cdot}9\times10^4$	17·5	11·6
$\frac{3}{8}$-inch smooth $x=1631$ cm	222	$1{\cdot}9\times10^4$	17·5	10·0
$\frac{3}{8}$-inch smooth $x=1631$ cm	136	$1{\cdot}2\times10^4$	16·1	12·8
$\frac{3}{8}$-inch rough $x=245$ cm	146	$1{\cdot}3\times10^4$	6·73	10·5
$\frac{3}{8}$-inch curved $x=250$ cm	113	$0{\cdot}9\times10^4$	15·0	21·9
$\frac{3}{8}$-inch curved $x=250$ cm	202	$1{\cdot}7\times10^4$	16·1	15·0

Note. The values of U/v_x for the curved pipe and for the rough pipe were measured. The values for the straight smooth pipe were taken from a curve representing previous experimental results at the appropriate value of R.

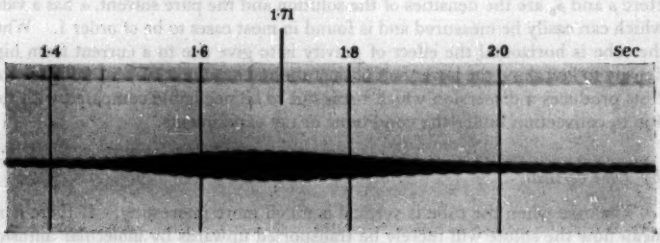


Figure 8.

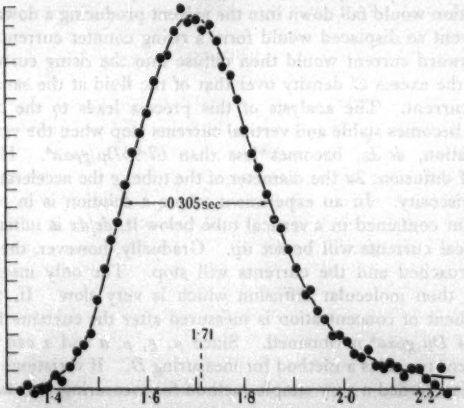


Figure 9. Dispersion in a very rough $\frac{3}{8}$-inch pipe.

The theory should apply to a rough pipe though not to a curved pipe. A pipe $2\frac{1}{2}$ metres long was artificially roughened by first wetting the inner wall with an adhesive and then pouring sand down it. Figures 8 and 9 show the record obtained and figure 9 the corresponding conductivity–time curve. Though the resistance coefficient was increased $6\frac{1}{2}$ times, the value of K/av_x, namely 10·5, was still close to the theoretical value 10·1. When the 0·9 cm pipe was bent into a circle about 3 feet diameter, K/av_x was increased about 50% at $R = 1{\cdot}7 \times 10^4$. These results are also given in table 2.

Effects of Gravity

(*a*) *Tube horizontal.*

In the preceding pages it has been tacitly assumed that the effect of gravity on the dispersion is negligible. To estimate the error involved in this erroneous assumption I calculated the dispersive power of gravity when the small variation of density due to dissolved material of concentration c can be represented by the expression

$$\rho = \rho_0(1 + \alpha c). \qquad \ldots\ldots(16)$$

Here ρ and ρ_0 are the densities of the solution and the pure solvent, α has a value which can easily be measured and is found in most cases to be of order 1. When the tube is horizontal the effect of gravity is to give rise to a current from high density to low along the lower half of the tube and a counter current along the top. This produces a dispersion which turns out to be negligible compared with that due to convection under the conditions of my experiments.

(*b*) *Tube vertical.*

The case when the tube is vertical is much more interesting. If there is no mean flow the solute will merely be transported upwards by molecular diffusion when the solution is below a lighter solvent, but when the heavier solution is above the solvent the equilibrium might be expected to be unstable. The heavier solution would fall down into the solvent producing a downward current and the solvent so displaced would form a rising counter current. The solute in the downward current would then diffuse into the rising current of solvent and so lose the excess of density over that of the fluid at the same level which drives the current. The analysis of this process leads to the condition that equilibrium becomes stable and vertical currents stop when the vertical gradient of concentration, dc/dz, becomes less than $67{\cdot}94 D\mu/g\rho\alpha a^4$. Here D is the coefficient of diffusion, $2a$ the diameter of the tube, g the acceleration of gravity and μ the viscosity. In an experiment where a solution is in contact with a lighter solvent contained in a vertical tube below it, dc/dz is initially very large, so that vertical currents will be set up. Gradually, however, the critical value will be approached and the currents will stop. The only means of vertical transport is then molecular diffusion which is very slow. If, therefore, the vertical gradient of concentration is measured after the currents have stopped, a measure of $D\mu/g\rho\alpha a^4$ is obtained. Since μ, g, ρ, a and α can be measured, the experiment provides a method for measuring D. If variations of D, μ and α with c are disregarded a very simple method for measuring D is to observe how

far down the tube the solute penetrates. If this is represented by Z, the value of D is $g\rho\alpha a^4 c_0/67{\cdot}94\mu Z$. Using this method I have obtained values of D which lie within the ranges of previous measurements.

References

Allen, C. M., and Taylor, E. A., 1923, *Trans. Amer. Soc. Mech. Engrs*, **45**, 285.

Fürth, R., and Ullman, E., 1926, *Kolloidzschr.*, **41**, 307.

Griffiths, A., 1911, *Proc. Phys. Soc.*, **23**, 190.

Hull, D. E., and Kent, J. W., 1952, *Industr. Engng Chem.*, **44**, 2745.

Smith, S. S., and Schulze, R. K., 1948, *Petrol. Engr*, **19**, 94 ; **20**, 330.

Taylor, Sir Geoffrey, 1953, *Proc. Roy. Soc.* A, **219**, 186; 1954 a, *Ibid.*, **225**, 473; 1954 b, *Ibid.*, **223**, 446.

The Effect of Intense Magnetic Fields on Electroluminescent Powder Phosphors

By A. N. INCE
Department of Engineering, University of Cambridge

Communicated by C. W. Oatley; MS. received 11th August 1954

Abstract. The effect of intense magnetic fields on electroluminescent powder phosphors is investigated to furnish information which should be helpful in developing a theory of the mechanism of electroluminescence. No quenching effect was observed even for magnetic field intensities as high as $1{\cdot}3 \times 10^5$ oersteds. The significance of this result is discussed.

§ 1. Introduction

Electroluminescence can be defined as the production of light by direct excitation of a phosphor with an electric field. The earliest indication of the influence of a d.c. field on the light emitted by a phosphor was reported in 1920 by Gudden and Pohl (1920). It was, however, Destriau (1936) who, for the first time, studied extensively the effect of strong a.c. fields on specially prepared ZnS phosphors. A comprehensive summary of his investigations has been reported (Destriau 1947). To explain the mechanism of electroluminescence, Destriau assumes direct field ionization of electrons in bound states of activator centres by the applied electric field. In other words, the electric field supplies enough energy to the electrons which are bound to the luminescent centres to enable them to make jumps of about 3 ev into the conduction band, and their subsequent return to the ionized centres produces electroluminescence. This theory has been modified by Curie (1953) who assumes that electrons which have been field-excited into the conduction band from donor levels about 0·75 ev deep and which have initial free paths of the order of 10^{-6} cm are accelerated by the applied field to such energies that they can ionize the luminescent centres by impact. Both of these authors believe that electroluminescence is mainly a volume effect and that the acceleration process takes place in a local field which is roughly equal to the externally applied field, although Curie admits the possibility of some surface states in the microcrystals of the phosphor powder where field strengths higher than those in the body of the crystal can exist and where the donors can be field ionized for subsequent acceleration in the body of the crystal. Therefore, taking the local field responsible for the acceleration of the electrons as 40 000 v cm^{-1}, and knowing the depth of the centres, which is about 3 ev for a green phosphor, they have come to the conclusion that the length of the paths along which the electrons are accelerated is of the order of 10^{-5} cm. In order to check this result Destriau (1947) has suggested that if a sufficiently intense magnetic field perpendicular to the electric field were applied to the phosphor, some quenching of the emitted light

would be possible. The relationship between the light output B of the electroluminescent cell and the applied field strength E, according to Destriau, is

$$B = aE^2 \exp(-b/E) \qquad \ldots\ldots(1)$$

where

$$b = p + qH^2, \qquad \ldots\ldots(2)$$

H being the magnetic field inensity and a, p and q constants. Hence he predicted that a magnetic field intensity of the order of 10^5 oersteds would be capable of producing quenching of electroluminescence. The highest magnetic field Destriau used in his experiments was 60 000 oersteds, for which no quenching was observed.

In recent times, phosphors with efficiencies much greater than those used by Destriau have been produced. These phosphors luminesce brightly at field strengths less than 10 000 v cm^{-1}. The use of these efficient phosphors excited at low electric field intensities would therefore require less magnetic field for quenching. The author, using an efficient green electroluminescent powder excited at about 10 000 v cm^{-1}, employed a magnetic field intensity of $1{\cdot}3 \times 10^5$ oersteds but did not observe any quenching effect. In the following paragraphs the significance of this result will be discussed and the experimental set-up used will be described.

§ 2. Experimental Method and Apparatus

The experimental arrangement is shown in figure 1. The electroluminescent cell was made by allowing a known amount of phosphor powder, suspended in a solution of polystyrene in benzene, to settle on a conducting glass electrode of area about $\frac{3}{8}$ in^2. The second electrode, which was again a conducting glass,

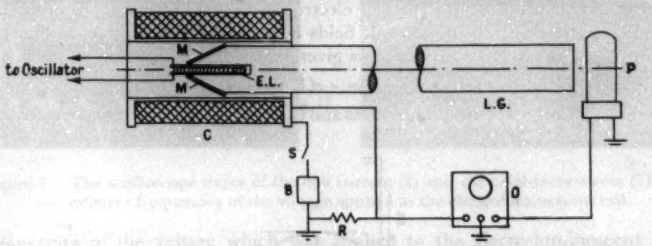


Figure 1. Experimental arrangement for the investigation of the effect of magnetic field on electroluminescent phosphors.

E.L., electroluminescent cell; L.G., Perspex light guide, $\frac{1}{2}$ in. diam. and 3 ft long; M, mirror; C, solenoidal coil, $\frac{5}{8}$ in. internal diam. and $1\frac{1}{2}$ in. long; P, photomultiplier; S, high speed switch; B, bank of condensers and power-pack; R, four-terminal resistance; O, oscilloscope.

was then stuck on top of the settled phosphor layer. In this way a cell with two transparent electrodes was obtained so that the light emitted from two surfaces was utilized. This cell was inserted into a slot made at the end of a Perspex light guide which had to be used in order to make possible the application of orthogonal electric and magnetic fields. The length of the light guide was made long enough so that the magnetic field did not affect the photomultiplier

which was used to pick up the light emitted by the cell. Two silver mirrors were placed on each side of the cell to increase the efficiency of light transmission through the guide. The output of the photomultiplier after being amplified by a d.c. amplifier was fed directly on to a pair of plates of an oscilloscope. To the second pair of plates was fed the voltage developed across a very small four-terminal resistance which was put in series with the solenoid coil. It was from the voltage developed across this resistance that the current and the magnetic field were computed.

The magnetic field was obtained by allowing a bank of condensers of 2500 μF charged to 2000 v to discharge through a suitably designed solenoidal coil (Ince 1954). The discharge from the condenser was oscillatory. The Q-factor of the circuit was about unity and therefore the stored energy was mainly dissipated during the first half cycle. The internal diameter of the coil was made as small as possible in order to increase the magnetic field for a given stored energy. The period of the oscillatory transient magnetic field was made considerably greater than the period of the brightness alternations to facilitate detection of any possible changes in the light output. Once the internal diameter and the period of oscillation have been settled all the coil parameters can be calculated (Ince 1954) to obtain the most efficient coil for a given stored energy.

§ 3. Theoretical Consideration

It can be shown (Shockley 1951) that an electron in a real crystal, under the influence of electric and magnetic fields, can be treated as a classical particle and consequently the laws of classical dynamics can, with a good approximation, be applied to it.

Let us consider, therefore, a free electron in vacuum placed at the origin and acted upon by electric and magnetic fields **E**, **H** whose directions are as shown in figure 2. The equation of motion is given by

$$\mathbf{F} = m\boldsymbol{\gamma} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{H}) \qquad \ldots\ldots(3)$$

where $-e$, m and v are the charge, mass and the velocity of the electron respectively.

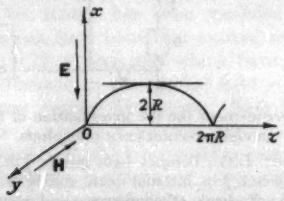


Figure 2. The cycloidal path of an electron in crossed electric and magnetic fields when the initial velocity is zero.

From equation (3) it can be shown (Millman and Seely 1941) that the increase of the kinetic energy of the electron, assuming the electron to be initially at rest, is given by

$$\tfrac{1}{2}mv^2 = 10^{16} m \left(\frac{E}{H}\right)^2 \sin\left(\frac{eH}{20m}\right) t. \qquad \ldots\ldots(4)$$

Hence $$(\tfrac{1}{2}mv^2)_{\max} = 10^{16} m(E/H)^2 \text{ ergs} \qquad \ldots\ldots(5)$$

where e is expressed in coulombs, m in grammes, E in v cm^{-1} and H in oersteds.

In the above calculations E and H have been assumed to be constant during the acceleration period of the electrons, which eventually excite luminescent centres by impact. This is justifiable since the period of acceleration of an electron to attain sufficient energy for electroluminescence is much shorter than the period of the externally applied alternating voltages. It has also been assumed that the electrons have zero initial velocities. A classical electron, moving in the crystal field, would be deflected so often that its vector velocity would take on practically random values and would, on the average, be zero.

In a real crystal the movement of the electron will be impeded by collisions, and its trajectory in crossed electric and magnetic fields will not therefore be a smooth cycloid but it will be composed of many arcs of a cycloid. Statistically this will cause the electron to gain less energy from the electric field than it otherwise would have done and, on the average, its maximum energy for given electric and magnetic fields will be less than that given by equation (5). Consequently the quenching effect of the magnetic field will be enhanced.

§ 4. Results and Discussion

The oscilloscope traces of the coil current, which is directly proportional to the magnetic field, and the brightness waves are shown in figure 3 for two different

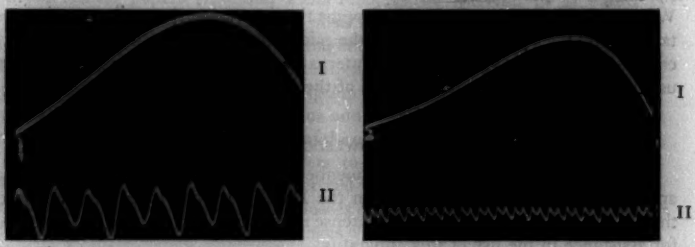


Figure 3. The oscilloscope traces of the coil current (I) and the brightness waves (II) for two different frequencies of the voltage applied to the electroluminescent cell.

frequencies of the voltage which was applied to the electroluminescent cell. The magnetic field had a maximum amplitude of 130 000 oersteds with a half period of 4 milliseconds. As can be seen from these oscillograms, there is no change either in the amplitude or in the wave-form of the brightness variations.

For the green electroluminescent phosphor used in these experiments the mean depth of the luminescent centres can be taken as 3 ev. The applied field strength E in the phosphor grains assumed to be homogeneous was approximately 10 000 v cm^{-1}, and was calculated from the expression (Roberts 1952)

$$E = E_{\mathrm{m}} \frac{3k_1}{2k_1 + k_2 - V_2(k_2 - k_1)}$$

where E_{m} is the field applied to the two-phase aggregate k_1, k_2 the dielectric constants of polystyrene and the phosphor powder respectively, and V_2 is the ratio of the phosphor volume to the total condenser volume.

Hence the minimum magnetic field intensity H, required for quenching, can be calculated from equatoin (5):

$$(\tfrac{1}{2}mv^2)_{\max} = 10^{16} m(E/H)^2 = 3 \text{ ev} = 3 \times 1{\cdot}6 \times 10^{-12} \text{ ergs.}$$

Putting $E = 10^4 \text{ v cm}^{-1}$ and $m = 9 \times 10^{-28}$ g, we obtain $H = 20000$ oersteds, whereas the applied magnetic field was about six times as great.

This result indicates that either the acceleration process of charges takes place in a region of the microcrystal where the field strength is at least, according to our results, ten times higher than in the rest of the crystal, so that the magnetic field strength required for quenching would be correspondingly higher, or the charges which are accelerated have a greater mass than the rest mass of an electron. Excluding the possibility of an ion excitation, which is difficult to accept, the second explanation of the experimental result does not seem to play any important part since velocities involved in the acceleration process are not so high as to make necessary the introduction into our calculations of relativistic mass correction. Of course the possibility cannot be dismissed that the assumptions upon which our calculations are based are not quite valid. For instance, the electronic energy which has been put equal to $\frac{1}{2}mv^2$ in our treatment may not be an isotropic function of the velocity (Curie, private communication). However, it is difficult to see how this anisotropy is going to create a condition under which a more efficient acceleration process than the one assumed in our calculations may be possible. On the other hand, there is separate experimental evidence (Piper and Williams 1952, Zalm *et al.* 1954) to support the theory that the acceleration process takes place in a region or regions of the microcrystals where the field strength is considerably higher than the mean internal field. These high field regions are usually associated with surface states of the crystal.

Acknowledgments

The author is indebted to Dr. Shoenberg, who kindly placed at his disposal apparatus which had been erected in the Royal Society Mond Laboratory for the production and measurement of intense magnetic fields, and to Mr. Laurmann for assistance in the use of this apparatus.

He also wishes to thank Thorn Electrical Industries Ltd., and the Research Laboratories of the General Electric Co. Ltd., for gifts of samples of electroluminescent powders, as well as the Admiralty for a grant towards the cost of apparatus.

References

Curie, D., 1952, *J. Phys. Radium*, **13**, 317; 1953, *Ibid.*, **14**, 672.
Destriau, G., 1936, *J. Chim. Phys.*, **33**, 587; 1947, *Phil. Mag.*, **38**, 700.
Gudden, G., and Pohl, R., 1920, *Z. Phys.*, **2**, 192.
Ince, A. N., 1954, *I.E.E. Monograph No.* 102. To be published shortly in *I.E.E. Proceedings*, Part IV.
Millman, J., and Seely, S., 1941, *Electronics* (New York : McGraw-Hill), p. 42.
Piper, W. W., and Williams, F. E., 1952, *Phys. Rev.*, **87**, 151.
Roberts, S., 1952, *J. Opt. Soc. Amer.*, **42**, 850.
Shockley, W., 1951, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).
Zalm, P., Diemer, G., and Klasens, H. A., 1954, *Philips Res. Rep.*, **9**, 81.

The Effect of Particle Shape Variations on the Coercivity of Iron Oxide Powders

By W. P. OSMOND†
Magnetic Tapes Division, Minnesota Mining and Manufacturing Co. Ltd., Slough, Bucks.

MS. received 30*th August* 1954

Abstract. A gaussian distribution function for the variation of the axial ratio m of powder particles is suggested as the explanation for the measured values of the coercivity of γ-ferric oxide powders being less than the theoretical values for particles of the observed mean shape factor $\bar{m}$. Superposition of the theoretical hysteresis loops for 20 packets of equal volume, each assumed to consist of particles with one value only of m, approximating to such a distribution gives values of bulk coercivity in very close agreement with measured values for two particular examples of widely different mean shape factors. In the light of these results explanations are offered of some other observed values for this oxide, and it is suggested that this mechanism may account in part for results obtained by Weil some years ago for precipitated ferronickels.

§1. Introduction

In two earlier papers (Osmond 1952 1953, to be referred to as I and II respectively) it was shown that the observed values of the intrinsic coercivity ${}_IH_c$ for dispersed fine powders of magnetic iron oxides (Fe_3O_4 and γ-Fe_2O_3), of the type commonly used for the magnetic coating of sound recording tapes, could be satisfactorily explained on the assumption that the powder particles were all of single ferromagnetic domain dimensions and of random orientation (Stoner and Wohlfarth 1948). Moreover, the ratio of the observed coercivities of many samples of quasi-acicular γ-ferric oxide powders to those of the respective magnetite powders from which they had been derived was shown in II to be closely proportional to the ratio of the respective losses of volume occurring in the course of the various chemical and structural changes which take place during the formation of these oxides from synthetic goethite (FeO.OH) pigment oxides, due allowance being made for the different values of saturation magnetic intensity.

To this extent, therefore, it was felt that these measured coercivities (much lower than theory would otherwise predict) provided support for the formula given by Néel (1947 b) denoting that the coercivity of a ferromagnetic substance consisting of single-domain particles separated by a fractional volume V of cavities should be reduced by the factor V, the evidence in this instance being that the internal cavities of the oxide particles provided this effect, the influence of the inter-particle spacing being negligible. However, a further point stressed by Néel in the same paper, namely the effect of the variation of shape which must be found among the particles of any powder, was not dealt with in either I or II, all measurements being referred to a mean shape factor (axial ratio of the equivalent prolate spheroid) which will henceforward be referred to as $\bar{m}$.

Study of this particular problem had actually been carried out some months before the publication of II, but presentation of the results in suitable form has been subjected to various delays. However, the appearance of a recent note on

† Now at Mullard Research Laboratories, Salfords, near Redhill, Surrey.

this subject by Wohlfarth (1954) has provided a new stimulus, and the results obtained are now given in this paper.

As in earlier work, throughout this paper measured and calculated field strength will be expressed in oersteds and magnetic intensity in gauss.

§2. Distribution of Shape Factors

The fact that the individual particles of an assemblage are most unlikely to have identical shape factors is probably the most important cause of the departures of observed hysteresis loops from the theoretical curve for identical particles of random orientation (Stoner and Wohlfarth 1948). Other factors may be (*a*) some slight departure from strictly random orientation throughout the aggregate and (*b*) some variations from stoichiometric chemical composition in the case of the oxides studied here. In the absence of attempts to obtain preferred orientation the former is likely to have little effect, while, with care in the various manufacturing processes, the latter (equivalent to variations in saturation magnetization I_0) should also be relatively unimportant.

Adopting with Wohlfarth (1954) the expression $g(m)\,dm$ for the fractional volume of the powder for which the shape factors of the particles lie between m and $m+dm$, the formula given by Néel (1947 b) is equivalent to a constant value of $g(m)$ so that in the whole range of possible values of m for prolate spheroids, from 1 (for the sphere) to infinity, the scales are heavily weighted in favour of extremely long, thin particles. However, Néel also used so rough an approximation for the expression for the demagnetizing factor $N(m)$ that, as Wohlfarth (1954) points out, it ceases to be of great accuracy for fairly low values of m and fixes the upper limit of the latter at 9/4. This would not be of great moment when applied to micropowder substances formed by the compression of metal particles derived by precipitation methods, of which the individual particles are likely to differ little from spheres, but the formula is clearly not suitable for quasi-acicular magnetic oxide particles with $\overline{m} \simeq 8$.

Wohlfarth (1954) has studied the effect of using a gaussian distribution function for $g(m)\,dm$, with particular reference to the fact that in ferronickels formed by precipitation methods Weil (1948) found values of ${}_IH_c/I_0 \sim 0{\cdot}5$, after extrapolation to allow for cavities, whereas Néel's formula should give 0·82. Since in the form used the distribution is about a mean value $\overline{m}=1$, Wohlfarth's formula is also unsuitable for $\overline{m} \simeq 8$.

The basic idea underlying the study described in the paper is that, in the absence of any other evidence, the most probable distribution is a gaussian function about the apparent mean shape factor $\overline{m}$ as estimated from electron-microphotographs. The function $g(m)$ then takes the form :

$$g(m) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{\frac{-(m-\overline{m})^2}{2\sigma^2}\right\} \qquad \ldots\ldots(1)$$

where σ is the standard deviation. By so choosing the value of σ that only a negligible fraction consists of particles with values of m as low as 1 we ensure that the idealized ellipsoids of Stoner and Wohlfarth's theory are all prolate. This of course restricts the effective upper limit for m to the value $2\overline{m}-1$, but when on photographic evidence the great majority of observable particles have values of m close to $\overline{m}$ this restriction seems unlikely to be a serious matter. From standard

tables for the normal probability curve (e.g. Attwood 1951), 0·99994 of the total volume will have values of m between 1 and $2\bar{m}-1$ if $\sigma=(\bar{m}-1)/4$, and this is the relationship adopted in the two examples described in the next section.

§3. Application to γ-Ferric Oxide Powders

3.1. *Non-Porous Particles with* $\bar{m}=1{\cdot}3$

The initial stimulus to this study was provided by the examination of a powder of γ-ferric oxide derived by the reduction and subsequent reoxidation of a red iron oxide powder described by the manufacturers as anhydrous Fe_2O_3. It was understood that the powder particles were grown in this form directly, and not by the dehydration of yellow FeO.OH oxides. It was therefore reasonable to assume that each particle consisted essentially of homogeneous α-Fe_2O_3. Moreover, the loss of oxygen on reduction to Fe_3O_4, subsequently regained on re-oxidation to γ-Fe_2O_3 would not be expected to cause the breaking up of each particle into smaller crystallites, in the manner found by Ervin (1952) on dehydrating α-AlO.OH (as described in II). Consequently the final magnetic powder could be considered to consist of homogeneous particles of γ-ferric oxide, for which it was shown in I that a suitable value of the saturation magnetization I_0 at room temperature is 400. Electron-microphotographs of the original red oxide showed that the individual particles were apparently all of shape factors lying between 1·2 and 1·4, and experience shows that no appreciable variation is to be expected during the subsequent chemical and structural changes. Adopting the mean value $\bar{m}=1{\cdot}3$, for which the demagnetizing factor $N(m)$ (i.e. the difference N_b-N_a between the demagnetizing factors along the equatorial and polar axes of the equivalent prolate spheroid) is 1·261 (Stoner and Wohlfarth 1948), the coercivity of an assemblage of identical particles of random orientation is 242, calculated by the formula

$$_IH_c=0{\cdot}48\,N(m)\,I_0 \qquad \ldots\ldots(2)$$

given by Stoner and Wohlfarth (1947, 1948) and by Néel (1947 b). The actual measured value for the powder was 210 approximately, while the hysteresis loop observed on a cathode-ray oscilloscope was, as usual, less abrupt in its steep portion than the theoretical curve.

It was clear that the results of a symmetrical distribution of values of shape factor m about $\bar{m}$ should lead to a reduction in the bulk coercivity, owing to the more rapid change of $N(m)$ with m as the latter decreases in value (cf. Stoner and Wohlfarth 1948, figure 8), and the most obvious natural distribution to assume was the gaussian form given by equation (1). Shunning the mathematics involved with this distribution in integrating Stoner and Wohlfarth's expression for the variation of resolved magnetization round the hysteresis cycle, it was decided to superpose the effects of 20 packets of equal volume, each being considered to consist of particles of random orientation (and negligible interactions) having one particular value of m. These 20 values of m were calculated from the values of x/σ in the standard table of areas under the normal probability curve corresponding to 0·025, 0·075, 0·125 . . . 0·925, 0·975 of the total, with the substitution of $|m-\bar{m}|$ for x. This approximation is of course a grouped distribution of the probabilities of occurrence, but it was felt that a smooth curve drawn through the summed effects of the 20 packets should give a very close approximation to that derived from a continuous distribution.

The value 242 found from equation (2) being so close to the extremely convenient figure $240 = 0{\cdot}48 \times 500$, it was decided to assume the latter as the coercivity of identical particles of the mean shape factor $\overline{m} = 1{\cdot}3$, an assumption which entails a very small reduction in the value of I_0. Hence, for these identical particles, Stoner and Wohlfarth's reduced field $\overline{h} = H/[N(\overline{m})\, I_0]$ becomes $\overline{h} = H/500$, so that $\overline{h}$ changes by 0·1 with every 50 oersteds change in H, a convenient maximum value of the latter being 1000 (i.e. $\overline{h} = 2$). For any packet of shape factor m the value of reduced field h is $H/[N(m)\, I_0]$, so that for every value of H the ratio of $h/\overline{h} = N(\overline{m})/N(m)$.

Adopting the relationship $\sigma = (\overline{m} - 1)/4$ mentioned at the end of § 2, we find that $\sigma = 0{\cdot}075$. This means that, with a smooth distribution, 50% of the total volume has shape factors in the range given by $x = \pm 0{\cdot}6745\sigma = \pm 0{\cdot}0506$, i.e. from $m = 1{\cdot}2494$ to $m = 1{\cdot}3506$. The 'coefficient of variation', defined as σ expressed as a percentage of $\overline{m}$, is here $7{\cdot}5/1{\cdot}3 = 5{\cdot}77$.

Table 1 gives the details for the 20 packets, the value of $h/\overline{h} = N(\overline{m})/N(m)$ being calculated from a large scale drawing of the variation of $N(m)$ with m, using the demagnetizing factors given by Stoner (1945).

Table 1

$\lvert x/\sigma \rvert$	1·9600	1·4395	1·1503	0·9346	0·7538	0·5978	0·4538	0·3186	0·1891	0·0628
$\lvert x \rvert$	0·147	0·108	0·086	0·070	0·057	0·045	0·034	0·024	0·014	0·005
Packet	1	2	3	4	5	6	7	8	9	10
m	1·153	1·192	1·214	1·230	1·243	1·255	1·266	1·276	1·286	1·295
$h/\overline{h}$	1·801	1·467	1·327	1·248	1·195	1·147	1·106	1·069	1·038	1·013
Packet	20	19	18	17	16	15	14	13	12	11
m	1·447	1·408	1·386	1·370	1·357	1·345	1·334	1·324	1·314	1·305
$h/\overline{h}$	0·725	0·778	0·814	0·841	0·864	0·888	0·910	0·934	0·963	0·989

The course of the calculation is then tedious but straightforward. For every value of H, varied by steps of 50 oersteds from 0 to 1000 and back to 0 (for the ascending and descending branches of the hysteresis loop), the contribution of each packet to the resolved magnetization I is found for its own particular value of h from a large scale drawing of the loop given by Stoner and Wohlfarth (1948) as amended by Rhodes (1948). The arithmetic mean of the 20 contributions is then the bulk value of I. The hysteresis loop obtained is shown in figure 1 together with Stoner and Wohlfarth's curve with ${}_IH_c = 240$ for comparison.

From this curve the bulk coercivity is read as 205 which compares admirably with the measured value of approximately 210 for the powder which gave rise to this study. The loop is less steep than the theoretical curve and the tips are more open, in general agreement with normal observation.

3.2. *Porous Particles with* $\overline{m} = 8$

It was next decided to carry out a similar calculation for a powder comparable with the quasi-acicular oxides in common use for sound-recording tapes, which exhibit coercivities of about 230 and consist of particles having shape factors, on the evidence of electron-microphotographs, mainly in the range 6 to 10. It was shown in II that the relatively low coercivity of these powders could be explained by the reduction of the effective demagnetizing factor $N(m)$ by a cavity factor V proportional to, but probably slightly less than, the fractional volume lost during

the dehydration of the original synthetic pigment oxide (α-FeO.OH), on the lines of the 'apparent density' formula

$$_{I}H_{c}=0{\cdot}48\,V\,N(m)\,I_{0} \qquad \ldots\ldots(3)$$

given by Néel (1947 b) for such structures containing cavities.

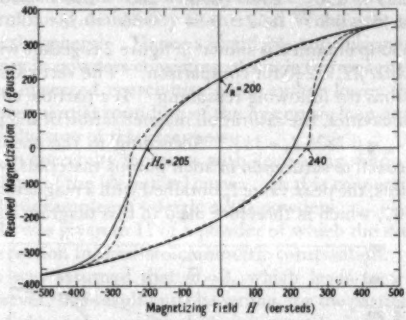


Figure 1. Full line, synthetic hysteresis loop for a powder of solid particles of γ–Fe_2O_3 of random orientation with a gaussian volume distribution of shape factors m about a mean value $\overline{m}=1{\cdot}3$. Broken line, theoretical loop for identical particles, all with $m=1{\cdot}3$.

With $\overline{m}=8$, for which $N(m)=5{\cdot}746$, $I_0=400$, and again taking 240 as the extremely convenient value of the coercivity for identical particles, we find $V=0{\cdot}2175$, which is quite an acceptable value in view of the estimated total loss of volume which was calculated in II as 0·235 approximately.

Using the same relationship between σ and $\overline{m}$ as before, we now have $\sigma=7/4=1{\cdot}75$, so that with a smooth distribution 50% of the total volume has shape factors in the range given by $x=\pm 0{\cdot}6745\sigma=\pm 1{\cdot}1804$, i.e. from $m=6{\cdot}8196$ to $m=9{\cdot}1804$. In this example the 'coefficient of variation' $=175/8=21{\cdot}87$. Table 2 gives the details for the 20 packets in this case, the values of $|x/\sigma|$ being the same as before.

Table 2

$\lvert x\rvert$	3·430	2·519	2·013	1·636	1·319	1·046	0·794	0·558	0·331	0·110
Packet	1	2	3	4	5	6	7	8	9	10
m	4·570	5·481	5·987	6·364	6·681	6·954	7·206	7·442	7·669	7·890
$h/\overline{h}$	1·127	1·072	1·051	1·039	1·028	1·022	1·016	1·010	1·006	1·002
Packet	20	19	18	17	16	15	14	13	12	11
m	11·430	10·519	10·013	9·363	9·319	9·046	8·794	8·558	8·331	8·110
$h/\overline{h}$	0·962	0·969	0·974	0·978	0·981	0·984	0·987	0·991	0·994	0·998

Comparison of these values of $h/\overline{h}$ with those given in table 1 shows that in this example all the values of h are much closer to $\overline{h}$ than in the comparable groups for the powder of smaller mean shape factor. This is due to the flattening of the $[N(m), m]$ curve in the region covering the relatively much larger values of m.

It could therefore be inferred that, not only should the bulk coercivity show a much smaller reduction from 240, but also the steep part of the hysteresis loop should be much closer to the abrupt rise of the curve for identical particles. For this reason the interval of 50 oersteds used before between successive values of H was lowered over the range 200 to 350, the actual values used being as follows :

H	200	225	250	260	270	280	290	300	325	350
h	0·4	0·45	0·5	0·52	0·54	0·56	0·58	0·6	0·65	0·7

The hysteresis loop obtained is shown in figure 2 together with the Stoner and Wohlfarth curve with ${}_IH_c = 240$ for comparison. The vertical scale for I in these curves is derived from the following reasoning. If a particle contains a fractional volume = 0·2175 of cavities, the saturation magnetization of the particle as a whole will be given by $I_m = 0{\cdot}7825\, I_0 = 313$. According to the theory given by Néel (1948) for the approach to saturation of such porous materials, which was applied in II to these powders, the peak value I_p obtained with a magnetizing field $H = 1000$ will be $0{\cdot}96\, I_m = 301$, which is therefore used in this diagram in place of $I_0 = 400$.

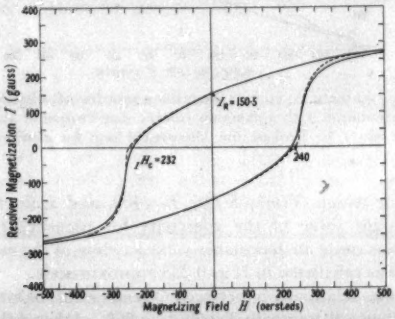


Figure 2. Full line, synthetic hysteresis loop for a powder of porous particles of γ–Fe_2O_3 of random orientation and effective porosity 0·2175 with a gaussian volume distribution of shape factors m about a mean value $\overline{m} = 8$. Broken line, theoretical loop for identical particles of the same basic coercivity (${}_IH_c = 240$) and peak magnetization ($I_p = 301$).

The further application of Néel's formula when the reduced field h is increasing above the critical value 0·5, which leads to the slight opening of the extreme tips of the loop as drawn, will not be described here, since it has negligible effect upon the coercivity which is the property studied in this paper.

From the loop the coercivity is read as 232 which again is in close agreement with measured values for such powders, usually ranging from 225 to 235. The reduction from 240, the basic value for identical particles is only 3·3%, compared with 14·6% in the first example, in spite of the fact that the 'coefficient of variation' is nearly four times as large ; moreover, the steep part of the loop lies much closer to the abrupt rise of the theoretical curve. Both these effects accord with the results expected from study of table 2.

§4. Discussion of Results

The two examples studied in the last section show that the assumption of a gaussian distribution of particle shape factors about an observed mean value can

satisfactorily explain the difference between measured values of the coercivity of dispersed magnetic oxide powders and the theoretical values for assemblages of identical particles. Although, for purposes of comparison with each other and with real powders, final values of magnetic field and intensity have been expressed in practical units (an extremely convenient conversion factor for the former being ready to hand), the calculations were actually carried out with the non-dimensional reduced units h and $\cos\phi$ defined by Stoner and Wohlfarth, so that the results obtained are perfectly general. Hence a justifiable conclusion is that variation of particle shape factors in powders consisting of single ferromagnetic domains is the major cause of the observed coercivities being rather lower than theory would predict, with a consequential rounding of the hysteresis loop at the bottom of the steepest part and reduction of this steepness.

The reduction in coercivity increases with decreasing shape factor on account of the increasing slope of the $[N(m), m]$ curve. In this connection it is of interest to touch on two other examples of γ-ferric oxide powders.

(*a*) An example was given in II of a powder of which the measured coercivity was 214·4 after correction for non-stoichiometric composition. In the absence of direct evidence it was assumed that $\overline{m}=8$, which leads to a 'porosity' factor $V=0{\cdot}1939$. However, this result must be wrong, for the particles could not all be identical. In the light of the present study it is more logical to assume a gaussian distribution function about a value of $\overline{m}$ somewhat less than 8. For the latter value we have found the reduction of coercivity to be 3·3%, and with $\overline{m}$ as low as 1·3 we have found 14·6%, while in another fully worked example with $\overline{m}=5$ the reduction was found to be 7·5% approx. Let us then suppose a reduction of 6% for this powder. Using equation (3) we have for identical particles: ${}_IH_c=192VN(\overline{m})=214/0{\cdot}94=228{\cdot}1$, a value which is satisfied by $\overline{m}=6$ (i.e. $N(\overline{m})=5{\cdot}468$), $V=0{\cdot}2173$, so that this example can be fitted into the same picture as that of § 3.2 without undue strain.

(*b*) In a recent paper by Westmijze (1953, p. 255) the measured coercivity of a magnetic tape coated with almost spherical particles is given as 75 (after conversion from m.k.s. units). Now this is identical with the value given by the formula

$$ {}_IH_c=0{\cdot}64\,K/I_0 \qquad \ldots\ldots(4)$$

for spherical particles of random orientation for which magneto-crystalline anisotropy is of paramount importance (Néel 1947a), with $I_0=400$ and $K=4{\cdot}7\times10^4$ (Birks 1950), although Westmijze did not obtain this agreement since, not knowing Birks' results, he used a much larger empirical value of K as well as a somewhat smaller value of I_0. However, without postulating a rigid boundary between the realms ruled by the two kinds of anisotropy, it is possible that the mean shape factor was sufficiently greater than 1 for the latter type of anisotropy to be of significance. With a gaussian distribution of such small values of m we could expect a value of bulk coercivity at least 15% below that for m in the light of the present study. Hence ${}_IH_c$ for $\overline{m}$ might be as much as 90, from which we find from equation (2) that $N(\overline{m})=0{\cdot}469$ which corresponds to a value of m very little less than 1·1.

One further point arising from the present study is worth noting. In the worked example of § 3.1, where no question of particle porosity arises and $\overline{m}=1{\cdot}3$, the calculated bulk coercivity is 205, the measured value for a comparable powder being practically the same at 210 approximately. With $I_0=400$ we have ${}_IH_c/I_0=0{\cdot}51$ or 0·52. Admittedly no allowance was made in this example for any

effect of inter-particle spacing, but the evidence of the calculations carried out in I and II is that this effect is negligible in these dispersed oxide powders. It therefore seems permissible to suggest that the above result may have some bearing upon the distribution of shape factors in the ferronickels studied by Weil (1948), referred to in § 2, in which shape anisotropy was considered to be the controlling factor and for which he found values of this ratio close to 0·5.

§ 5. Conclusion

Since this paper was first written it has been found possible to study a new article by Néel (1954), in which he enumerates the factors which must be taken into account in forming an adequate theory of the hysteresis exhibited by magnetically hard modern permanent magnets. As is now generally accepted, these substances consist essentially of single domain ferromagnetic grains surrounded by a non-magnetic medium, and are therefore comparable with the magnetic tape coatings studied here. Néel lists five important factors : (i) grain orientation, (ii) grain proximity (i.e. influence of cavities in the agglomerate), (iii) distribution of individual grain coercivities, (iv) inter-grain fields of a Lorentz type, (v) fluctuations of these fields, and by means of a series of admittedly very rough approximations he builds up a schematic theory of the ideal permanent magnet, taking account of all five factors.

In the present paper the first three of these factors are taken into account, and in as precise a way as possible, without the use of drastic simplifications. Néel estimated the effect on his model of the last two factors in weak fields, but owing to the mathematical complications involved at high magnetizations and in the remanence region he found it impossible to extend the study in a relatively short paper. For similar reasons these factors will not now be studied here. Important though they may well be, in particular as possibly affording an explanation of the observed values of remanent magnetization being usually somewhat less than half the saturation value (cf. I and II), this paper shows that the observed values of the coercivity of the dispersed γ-ferric oxide powders discussed here can be satisfactorily explained by careful consideration of the first three only of Néel's factors.

Acknowledgments

I am indebted to Mr. P. T. Hobson, Production Manager for Magnetic Tapes, and to the Directors of the Minnesota Mining and Manufacturing Co. Ltd., for permission to publish this paper, and I am grateful to the former for his interest in this study and to Dr. E. P. Wohlfarth for his encouragement.

References

Attwood, C., 1951, *Advanced Five-Figure Mathematical Tables* (London : Macmillan), p. 57.
Birks, J. B., 1950, *Proc. Phys. Soc.* B, **63**, 65.
Ervin, G., 1952, *Acta Cryst., Camb.*, **5**, 103.
Néel, L., 1947 a, *C.R. Acad. Sci., Paris*, **224**, 1488; 1947 b, *Ibid.*, **224**, 1550; 1948, *J. Phys. Radium*, **9**, 184; 1954, *Appl. Sci. Res., Hague* B, **4**, 13.
Osmond, W. P., 1952, *Proc. Phys. Soc.* B, **65**, 121; 1953, *Ibid.*, **66**, 265.
Rhodes, P., 1948, *Proc. Leeds Phil. Lit. Soc.*, **5**, 116.
Stoner, E. C., 1945, *Phil. Mag.*, **36**, 803.
Stoner, E. C., and Wohlfarth, E. P., 1947, *Nature, Lond.*, **160**, 650; 1948, *Phil. Trans. Roy. Soc.* A, **240**, 599.
Weil, L., 1948, *C.R. Acad. Sci., Paris*, **227**, 1327.
Westmijze, W. K., 1953, *Philips Res. Rep.*, **8**, 148–157, 161–183, 245–269, 346–366.
Wohlfarth, E. P., 1954, *Research*, **7**, S1.

On the Abrasion of Rubber

By A. SCHALLAMACH
The British Rubber Producers' Research Association, Welwyn Garden City, Herts.

MS. received 20th September 1954

Abstract. The following theoretical results on the abrasion of rubber have been deduced from a few simple assumptions concerning the initiation of the surface damage. The abrasion is proportional to the normal load, independent of the particle size of the abrasive if the particles are polyhedral, and proportional to their mean radius of curvature if they can be approximated to hemispheres. The spacing of the abrasion pattern is proportional to the cube root of the normal load, proportional to the two-thirds power of the particle size on an abrasive with polyhedral particles, and directly proportional to the particle size on an abrasive with hemispherical particles.

These predictions have been reasonably well confirmed by experiment.

§ 1. Introduction

In the course of investigations into the abrasion of rubber, model experiments have recently been carried out (Schallamach 1952 a, to be referred to as I) in which the conditions thought to obtain in the interface between rubber and abrasive were simulated. This was done by scratching rubber with a needle under controlled conditions, and the ensuing surface damage was found to depend markedly on the type of rubber compound employed. The general characteristics of the needle traces could be qualitatively explained in terms of the mechanical strength and the elastic and frictional properties of the samples, and various mechanisms were observed by which particles of rubber can be detached from the bulk under localized stress concentrations such as may be expected to occur where the proud particles of the abrasive impinge on the rubber surface. The needle traces gave necessarily an exaggerated picture of the extent of the surface damage produced in abrasion, but not necessarily a wrong picture of its mechanism. It will be shown here that similar results are obtained under milder conditions when the stress concentrations are brought about by friction; on the basis of this and a very few other assumptions, expressions for the dependence of abrasion on normal load and on the nature of the abrasive are deduced and compared with experimental results.

The same line of approach is shown to lead to an explanation of the origin and properties of the so-called abrasion pattern. This phenomenon has been described before (Schallamach 1952 b, to be referred to as II); it consists of a system of nearly parallel ridges at right angles to the direction of abrasion often found on abraded rubber surfaces. Theoretical relations connecting the spacing of the abrasion pattern with normal load and coarseness of the abrasive are now given and tested experimentally.

§ 2. The Model

When a sharp point is pressed on to a rubber surface and is set into tangential motion, a surface traction is produced on the rubber either because the point has punctured the surface and has become mechanically entangled with the rubber, or because of frictional adhesion. In the most severe of our experiments the load per abrasive particle was only 7 g, under which thrust even a needle will not penetrate the surface of rubber compounds of the type used. We attribute therefore the tangential stress initiating surface failure mainly to friction. The stress will be inhomogeneous and extremely complicated but a qualitative picture for the two-dimensional analogue can be obtained by means of a photo-elastic experiment. Figure 1 (Plate I) shows a cylindrical frictional glider of 30 mm diameter exercising the tangential force given by its limiting friction on one of the end faces of a sheet of transparent rubber 19 mm thick. The sample was photographed in sodium light between crossed Polaroid sheets and quarter-wave plates so that only isochromatics are visible. The isochromatics clearly indicate a stress concentration at the rear edge of the glider, and as the concentration occurs near the free surface, the rubber must be under tensile stress immediately behind the glider. If the rubber fails, tearing at right angles to the direction of motion would be expected. That this is so has been shown by means of an experiment carried out with the same apparatus with which the needle scratching experiments had been made, the needle having been replaced by a hemispherically ended wire of 1 mm diameter. The rubber compounds employed for this purpose were relatively soft in order to keep the normal loads within bounds (cf. equation (1)). Two typical traces are shown in figure 2 (Plate I) which gives photomicrographs of the surface damage on two unfilled vulcanizates of natural rubber. The traces are similar to some of the traces produced by a needle as shown in figure 3 of paper I.

Once a tear is opened, a particle will eventually be detached from the bulk during successive passes of the abrasive by any of the mechanisms discussed in detail in I. In order to arrive at theoretical expressions for the abrasion the following assumptions are made: (i) The length of the tear opened by the abrasive particle is proportional to the width of the area of contact between the abrasive particle and the rubber. This assumption is suggested by figure 2 where the tears appear to be confined in length to the width of the trace of the blunt point marked by a faint roughening of the rubber surface. (ii) The volume of the detached rubber particle is proportional to the third power of the length of the original tear, irrespective of the way in which the particle is finally removed.

If the mean diameter of the area of contact is a, the normal load W, the mean radius of curvature at the tip of the abrasive particle where it is in contact with the rubber r, and G an elastic constant characterizing the rubber, it follows from simple dimensional considerations that

$$a/r = \text{const.}\ (W/Gr^2)^{\alpha} \qquad \ldots\ldots(1)$$

where the exponent α has to be determined experimentally. The justification for introducing mean values for a and r, and for assuming that the elastic behaviour of the rubber can be described by a single constant will be given presently. If there are n^2 abrasive particles per unit area, and the normal load per unit area is L so that $W = L/n^2$ then the abrasion per unit area, A, should, according to assumptions (i) and (ii) above, be given by

$$A \propto n^2 a^3 = \text{const.}\ n^2 r^3 (L/Gn^2r^2)^{3\alpha}. \qquad \ldots\ldots(2)$$

The physical properties of the rubber enter equation (2) as G and as a factor, contained in the constant, which indicates how many passes of the abrasive are necessary for the detachment of a rubber particle, and which is therefore a measure of the abrasion resistance. This problem is not dealt with in the present paper. It must be pointed out that the elastic properties of loaded rubber compounds depend pronouncedly on previous deformations to which the samples have been subjected (Mullins 1947) and it has been shown (II; Schallamach 1953) that the top layer near an abraded rubber surface is in general much softer than the bulk material because of the repeated large deformations occurring there. The effective value of G in equation (2) is therefore different from that found with conventional measurements.

§ 3. The Area of Contact

In order to arrive at the value of α to be used in equations (1) and (2) recourse was had to a model experiment in which a large sample of rubber, 25 mm thick and 10 cm in diameter, was indented by a hemispherical indenter of 25 mm diameter. By ruling equidistant lines on the rubber and having a transparent indenter made of Perspex the diameter of the area of contact could be read off

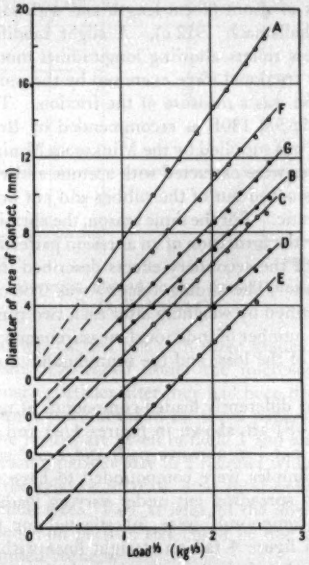


Figure 3. Diameter of area of contact plotted against cube root of normal load. For clarity, the curves are plotted from different origins.

directly. The results of these measurements on seven different compounds, the composition of which is given in the Appendix, are reproduced in figure 3 where the diameter of the area of contact is plotted against the cube root of the indenting load. It is seen that with this method of presenting the results straight lines are obtained which very nearly pass through the origin. The parameters

of the lines were obtained by the method of least squares. The relation between a and W may therefore be expressed to a good degree of approximation by the equation $a \propto W^{1/3}$. It is interesting to note that this equation is the same as that given by Hertz for materials having ordinary elasticity. An easily accessible account of Hertz's work can be found in Timoshenko's book (1934) where it is also shown that in the general case of a non-spherical indenter the area of contact is an ellipse the axes of which again obey the cube root law. This is taken as sufficient justification for having introduced in equations (1) and (2) the mean radius of curvature r and only one elastic constant, G.

§ 4. The Load Dependence of Abrasion

Introducing the exponent $\alpha = 1/3$ into equations (1) and (2) gives

$$a = \text{const.}\,(Wr/G)^{1/3} \qquad \ldots\ldots(3)$$

and

$$A = \text{const.}\,Lr/G. \qquad \ldots\ldots(4)$$

Equation (4) gives the simple result that the abrasion should be proportional to the normal load. The experimental test of this relation was carried out on a straight abrasive track of about 60 cm length and with samples similar to those described before (Schallamach 1952 c). A slight modification was to mount the track on frictionless rollers allowing longitudinal motion but the track was held back, against the frictional force exercised by the moving sample, by a leaf spring whose deflection was a measure of the friction. The abrasive employed was garnet paper grade 5/0 180E as recommended by British Standard Specification 903 : 1950 and was supplied by the Minnesota Mining and Manufacturing Company. All samples were extracted with acetone and vacuum dried as otherwise mobile impurities oozed out of the rubber and not only falsified the results but also made them erratic. For the same reason, the abrasive was only used once.

In order to prevent the formation of an abrasion pattern which would increase the abrasion because of the secondary effects described in II the samples were reversed after each run. The sliding velocity was 0·66 cm sec^{-1}, and the rate of abrasion was determined by weighing after each two runs, making an abrasive path of 128 cm. The number of individual measurements at each load depended on the absolute value of the loss, and the reproducibility of the data lay within $\pm 3{\cdot}5\%$.

The results for four differently loaded compounds of natural rubber and one of Krylene (Cold GR–S) are shown in figures 4(*a*) and 4(*b*), where they are plotted as loss in cm^3 per 1 m abrasive path and 1 cm^2 surface against normal pressure. All these samples were compounded to have similar hardnesses so that any effects due to spreading out under varying loads would be small and similar; no unfilled compounds were investigated for this reason. All the experimental points in figure 4 fall on straight lines within the limits of error but the lines when produced do not pass through the origin. The prediction of a linear relation between abrasion and normal load as given by equation (4) is therefore largely fulfilled, apart from a small divergence at low loads. The probable explanation of this discrepancy is that because of the irregularities of both surfaces new contacts are made when the load is increased from very low values, thus increasing the total area of contact above that given by equation (3). This would agree with the frictional behaviour to be described in § 6.

§ 5. The Influence of the Abrasive

According to equation (4) the abrasion should be proportional to the mean effective radius of curvature of the abrasive particles so that, when comparing abrasions on different grades of the same abrasive, the relation between particle size and radius of curvature ought to be known. There appears to be no *a priori* reason why the sharpness of the corners of polyhedral particles should depend on their size, and it is assumed that, to a first approximation, the radius of curvature is independent of size; the abrasion would therefore be expected to be independent of the grade of abrasive.

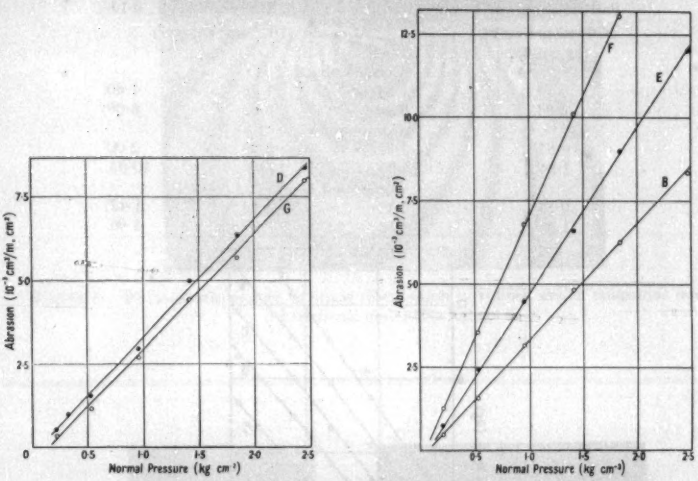


Figure 4 (*a*) and (*b*). Abrasion plotted against normal pressure of compounds B, D, E, F and G.

An experimental test was carried out with three grades of garnet paper, the particle size having been determined under the microscope by measuring a number of randomly chosen particles after they had been dissolved off the paper. Three different compounds were abraded on these papers under two different normal loads each. The results are given in table 1 and show that the abrasion does increase with increasing particle size in a manner which appears to depend on the compound but the relative increase is at most only 70% for an almost five-fold increase in particle size. Part, at least, of the observed departure from the predicted independence on particle size could be due to the larger particles having rather more rounded corners.

Abrasion experiments on a different type of surface will be discussed separately in § 8.

§ 6. Friction on an Abrasive

It has been shown (Schallamach 1952 c) that the frictional force of rubber is proportional to the true area of contact between rubber and track, and that on smooth glass the area of contact is proportional to $L^{2/3}$. It will be seen that

in this case the area of contact obeys an equation identical with equation (3). It has been stressed by Lodge and Howell (1954) that the frictional force of an elastic solid depends on the geometry of the track surface, this being in marked contrast to the friction of ordinary solids which is governed by their plastic flow properties (Bowden and Tabor 1950, King and Tabor 1954). If the frictional glider is a sphere large compared with the asperities of the elastic solid, it is easily shown on the basis of Hertz's theory (Lincoln 1953, Lodge and Howell 1954)

Table 1

Grade of abrasive	1 1/2 40E	1/0 80E	5/0 180E
Particle size (mm)	0·63	0·29	0·13
Normal pressure (kg cm^{-2})	Abrasion (10^{-3} cm^3/m, cm^3)		
	Compound D		
0·53	1·81	1·76	1·60
1·84	8·97	6·82	6·09
	Compound E		
0·53	3·23	2·97	2·00
1·84	13·70	12·05	10·05
	Compound G		
0·53	2·11	1·83	1·43
1·84	10·10	7·50	5·96

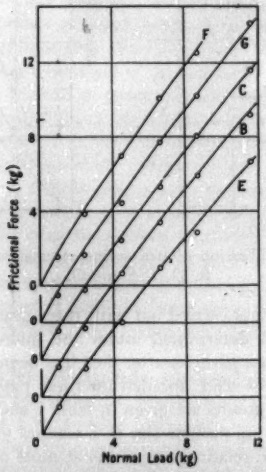


Figure 5. Frictional force plotted against normal load, with the following values of ϕ: B, 1·75; D, 1·765; E, 1·66; F, 1·86; G, 1·80.

that the true area of contact is proportional to $W^{8/9}$ where W is the load on the glider. According to our conception of the nature of the abrasive, this result can be applied to the frictional force F between rubber and abrasive:

$$F = \phi L^{8/9} \qquad \ldots\ldots(5)$$

where ϕ is a constant involving, among other things, the dimension of the asperities, elastic constant of the rubber and interfacial energy between rubber and track.

Our results are shown in figure 5 where the full lines have been calculated according to equation (5), with ϕ adjusted to give the best fit at high normal loads.

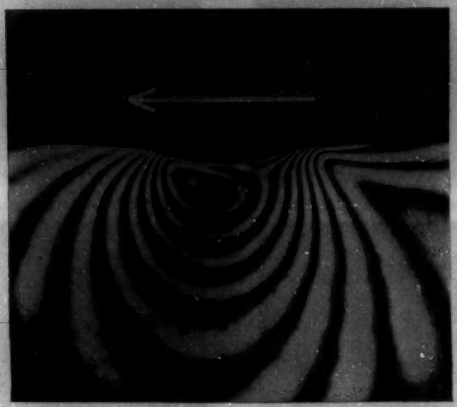
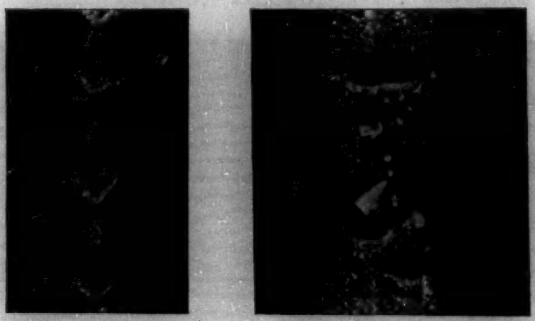


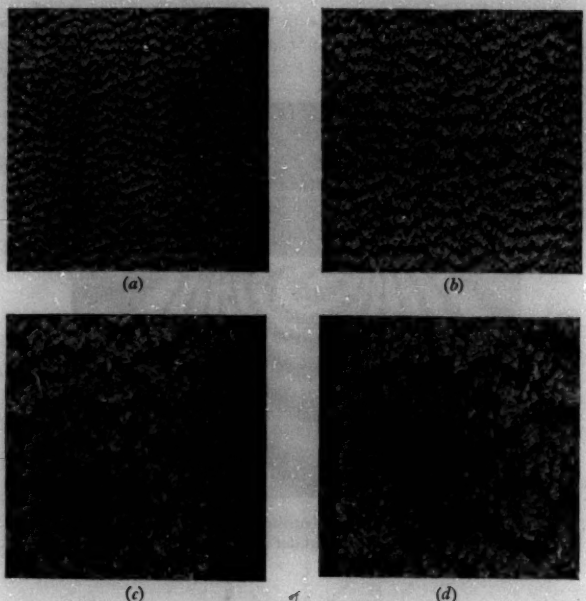
Figure 1. Photo-elastic picture of stress distribution in rubber under tangential traction of frictional glider (Na light).



(*a*) (*b*)

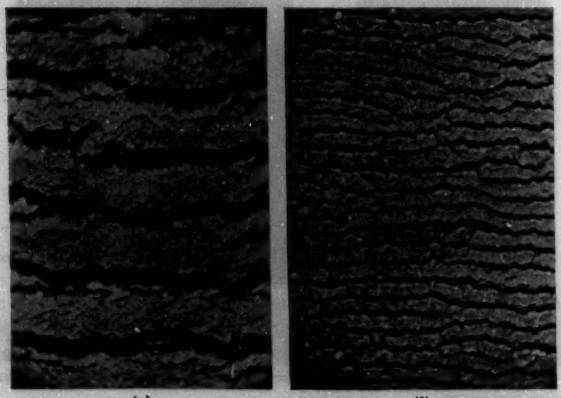
Figure 2. Traces of a blunt point of 1 mm diameter under a load of 80 g on unloaded natural rubber vulcanizates of (*a*), $M_{300}=12\,\mathrm{kg\,cm^{-2}}$ ($\times 18$) and (*b*), $M_{300}=21\,\mathrm{kg\,cm^{-2}}$ ($\times 40$). M_{300} = stress per original cross section at 300% elongation. Direction of motion downward.

PLATE I.



(*a*) (*b*) (*c*) (*d*)

Figure 6. Abrasion patterns on compound E. Normal pressures: (*a*) and (*c*) 0·53 kg cm^{-2}; (*b*) and (*d*), 1·84 kg cm^{-2}. Abrasives: (*a*) and (*b*), 0·13 mm; (*c*) and (*d*), 0·63 mm. (×14.)



(*a*) (*b*)

Figure 7. Abrasion patterns formed on concrete road surfaces. (*a*), coarse, and (*b*), fine surface. (×14.)

PLATE II.

The experimental points tend to fall somewhat below the theoretical curve at low normal loads. This is thought to be due to the same cause as the low abrasion values in this region discussed in § 4.

For many practical purposes, the coefficient of friction on an abrasive can be considered as constant so that, to the same degree of approximation, abrasion may be said to be proportional to the work done on the rubber. It is interesting to note that Powell and Gough (1954) advocate abrasion measurements at constant power input for the road wear assessment of tyre compounds.

§ 7. The Abrasion Pattern

The mechanism of rubber abrasion as envisaged in § 2 begins with the opening of a number of tears in the path of the abrasive particles which on subsequent passes of the abrasive are propagated in such a manner as to separate a rubber particle from the bulk. During this process, the front edge of the tear will be elongated in the direction of motion of the abrasive, similar to the thongs pulled out by a needle as shown in figure 5 of I. The lip formed in this way will cover one or more tears in front which will thus be protected from further damage; when the lip has snapped back the abrasive can attack another uncovered tear. A periodic structure will be produced on the surface which has the characteristics of the abrasion pattern as described in II. In particular, the structure will develop gradually and will be the more pronounced the softer the rubber.

If the abrasion pattern originates in this way, the dependence of the spacing of the ridges on normal load and on the grade of the abrasive can be estimated. The extent of the lip, which determines the spacing, must depend on the volume involved in the elementary abrasion process which in its turn has been assumed in § 2 to be proportional to a^3. For dimensional reasons, the spacing s should then be proportional to a so that from equation (3):

$$s \propto (Wr/G)^{1/3} \propto (Lr/n^2G)^{1/3}. \qquad \ldots\ldots(6)$$

According to this expression, s increases only with the cube root of the normal load; the properties of the abrasive enter equation (6) with r and also with n, i.e. the grade of the abrasive. For the present purpose the connection between n and the particle size d is given with sufficient accuracy by the equation $d = 1/n$, and the spacing should increase with $d^{2/3}$. As for the variation of r with particle size, this can be deduced from the observed increase of the abrasion with increasing particle size as given in § 5. The experiments described below have been carried out on the garnet papers 1 1/2 40E and 5/0 180E; the average ratio of the abrasions on these two papers is 1·445 which, if totally attributed to the increase of r, means, according to equation (4), that the effective radii of curvature have the same ratio. The relative increase of the spacing due to this reason would be, from equation (6), equal to $1{\cdot}445^{1/3}$, i.e. 1·10.

Abrasion patterns were produced on three different compounds at a velocity of 6·60 cm sec^{-1}. This velocity is higher than was used for the abrasion measurements because the patterns develop better at higher rates of sliding, the reason for this effect not being properly understood. On the coarse abrasive, the patterns tend to be obscured by heavy scoring marks which limit the scope of these experiments to a certain extent. The results are given in table 2 where the values of the spacings are averages of eight counts. Examples of abrasion patterns developed under various conditions are shown in figure 6 (Plate II). The magnification employed when photographing the coarse patterns is actually

too high for these surfaces but it is desirable to show all patterns at the same magnification. In table 2 r_L is the ratio of the spacings produced under the

Table 2

Particle size of abrasive (mm)		0·13	0·63	
Normal pressure (kg cm^{-2})		spacing (cm)		
		Compound C		r_a
0·53		0·0123	0·0398	3·23
1·84		0·0183	—	
	r_L	1·49		
		Compound D		
0·53		0·0145	0·0470	3·24
1·84		0·0191	—	
	r_L	1·315		
		Compound E		
0·53		0·0147	0·0508	3·45
1·84		0·0211	0·0795	3·75
	r_L	1·435	1·565	

loads of 1·84 and 0·53 kg cm^{-2}. The theoretical value of r_L is $(1{\cdot}84/0{\cdot}53)^{1/3} = 1{\cdot}515$. r_a is the ratio of the spacings produced on the two abrasives, the theoretical value of which is $1{\cdot}10(0{\cdot}63/0{\cdot}13)^{2/3} = 3{\cdot}10$. When comparing the theoretical ratios with those given in table 2 it will be seen that the theory is at least qualitatively borne out by experiment.

§ 8. Abrasion on Road Surfaces

If the abrasive can be approximated to a set of close-packed hemispheres, so that $r = 1/n$ in equation (3), it follows from equations (4) and (6) that

$$A/s = (L/G)^{2/3}. \qquad \ldots\ldots(7)$$

Abrasion and spacing of the pattern are therefore proportional to each other with this type of abrasive. However, the value of A in equation (7) is the abrasion determined in the absence of a disturbing abrasion pattern, and in principle A and s should be determined in separate experiments.

A set of hemispheres can perhaps serve as a model of a concrete surface consisting of sand or gravel embedded in cement. A few abrasion measurements have been carried out on two experimental concrete road surfaces in the grounds of the Road Research Laboratories at Harmondsworth. These surfaces cannot be specified in quantitative terms, and they will simply be described as very coarse and very fine. Only one compound, C, was abraded on both surfaces, and it was not practicable to reverse the direction of abrasion during the experiments. As a consequence, pronounced abrasion patterns were formed on both surfaces which are reproduced in figure 7 (Plate II). The values of the abrasions under a normal pressure of 1·65 kg cm^{-2} and at a sliding velocity of 48 cm sec^{-1}, and the lengths of the spacings produced at the same time are given in table 3.

Table 3

Surface	Coarse	Fine	Ratio
Abrasion (10^{-3} cm^3/m, cm^2)	1·57	0·41	3·88
spacing (cm)	0·832	0·234	3·56

The good agreement between the ratios of abrasions and spacings shown in table 3 is to a certain extent misleading, as the abrasion patterns will have increased

the abrasions on the two concrete surfaces to different extents. But even if this effect is allowed for by reducing the abrasion ratio by, say, 20% there is still substantial agreement with equation (7), and comparison of the results obtained on garnet papers of different grades with the results in table 3 demonstrates the important bearing of the nature of the abrasive on the laws of abrasion.

§ 9. Conclusions

The most sweeping assumption made in our theoretical approach to the abrasion problem is that all rubbers are subject to the same abrasion mechanism irrespective of polymer and filler. Nevertheless, it will have been seen that the naïve conceptions put forward in § 2 lead to predictions which are at least qualitatively borne out by the experimental findings. It would appear therefore that the assumptions which have been made are essentially correct and thus provide a basis for a quantitative approach.

Acknowledgments

The author wishes to thank the Director of the Road Research Laboratory and Mr. A. G. Tarrant for having made possible the experiments referred to in § 8.

This work forms part of a programme of research undertaken by the Board of the British Rubber Producers' Research Association.

References

Bowden, F. P., and Tabor, D., 1950, *The Friction and Lubrication of Solids* (Oxford: Clarendon Press).

King, R. F., and Tabor, D., 1954, *Proc. Roy. Soc.* A, **223,** 225.

Lincoln, B., 1953, *Nature, Lond.*, **172,** 169.

Lodge, A. S., and Howell, H. G., 1954, *Proc. Phys. Soc.* B, **67,** 89.

Mullins, L., 1947, *J. Rubb. Res.*, **16,** 275.

Powell, E. F., and Gough, S. W., 1954, *Proc. Third Rubb. Technol. Conf.*, in the press.

Schallamach, A., 1952 a, *J. Polym. Sci.*, **9,** 385 (I); 1952 b, *I.R.I. Trans.*, **28,** 256 (II); 1952 c, *Proc. Phys. Soc.* B, **65,** 657; 1953, *Ibid.*, **66,** 817.

Timoshenko, S., 1934, *Theory of Elasticity* (New York and London: McGraw-Hill).

APPENDIX

Composition of Samples

	A	B	C	D	E	F	G
Smoked Sheet	100·0	100·0	100·0	100·0	100·0	100·0	—
Krylene	—	—	—	—	—	—	100·0
Zinc Oxide	5·0	5·0	5·0	5·0	5·0	5·0	5·0
Sulphur	2·5	3·0	2·5	2·5	2·5	2·5	1·8
Stearic Acid	1·0	3·0	2·5	3·0	3·0	1·0	3·0
Dutrex R	—	—	—	2·0	2·0	—	2·0
Pine Tar	—	—	4·5	—	—	—	—
Santocure	0·6	—	0·5	0·7	0·7	0·7	1·1
M.B.T.	—	1·0	—	—	—	—	—
P.B.N.	1·0	1·0	1·0	—	—	—	—
Nonox H.F.N.	—	—	—	1·0	1·0	1·0	1·0
Filler	—	50 MPC	50 HAF	50 HAF	75 FT	130 W	50 HAF
Cure, (min at 140°c)	50	50	40	35	40	20	80 at 150°c

Notes and abbreviations: Krylene is a low-temperature polymerized co-polymer of butadiene and styrene; Dutrex R is a mineral oil; M.B.T. (Mercaptobenzothiazole) and Santocure are accelerators; P.B.N. (Phenyl-β-naphthylamine) and Nonox H.F.N. are antioxidants; HAF=high abrasion furnace black; MPC=medium processing channel black; FT=fine thermal black; W=Pattinson's activated $CaCO_3$ (whiting).

RESEARCH NOTES

Electrical Conductivity of Some Condensed Aromatic Hydrocarbons

By D. C. NORTHROP and O. SIMPSON
Services Electronics Research Laboratory, Baldock, Hertfordshire

MS. received 11*th August* 1954

ELECTRICAL conductivity and photoconductivity in condensed aromatic hydrocarbons have been reported by Mette and Pick (1953) and by Inokuchi (1954). The same materials are fluorescent, and solid solutions of one in another show the phenomenon of fluorescence transfer from solvent to solute in the manner described by Bowen (1944) and by Simpson and Northrop (1954). Investigation of solid solutions promises to throw considerable light on the properties of these hydrocarbons, and the purpose of this communication is to describe measurements of the electrical conductivity of four pure hydrocarbons and some solid solutions. We have measured the temperature dependence of specific conductivity of anthracene, naphthacene, pentacene and perylene, and solid solutions in anthracene of the latter three materials, in concentrations 10^{-3} and 10^{-4} molar. All the solid solutions showed fluorescence transfer.

The hydrocarbons have very low conductivities, for example that of anthracene is 5×10^{-18} ohm^{-1} cm^{-1} at room temperature, and measurements have therefore been made on thin sandwich-type specimens vacuum deposited on Pyrex flats. The pure hydrocarbons were evaporated from a nickel gun, and the solid solutions by a co-evaporation technique. Electrical contact was made by means of evaporated lead electrodes, the bottom one evaporated directly on the Pyrex, and the upper one entirely on top of the hydrocarbon (see figure 1). The overlap of the two electrodes defined a square of 0·4 cm^2, and the thicknesses of the hydrocarbon films varied between 2×10^{-4} cm and 5×10^{-4} cm. Each specimen was enclosed in a copper calorimeter whose temperature could be varied and which acted as an electrostatic screen. A 1·5 volt battery was connected in series with the specimen and the current measured by means of a balanced pair of electrometer triodes whose current sensitivity was 10^{-14} amp. An equilibrium current was reached only after a long period, but under steady state conditions the specimens are ohmic at the temperatures and applied fields used in these experiments. The temperatures lay always within the range 0°c to 100°c, the lower limit being decided by the sensitivity of the current measuring circuit, and the upper limit by various forms of damage to which the specimens proved liable on heating.

Throughout the range of measurements the pure materials were found to obey an exponential dependence of specific conductivity on the reciprocal of the absolute temperature $\sigma = \sigma_0 \exp(-\epsilon_0/2kT)$ where σ_0 and ϵ_0 are constants characteristic of individual hydrocarbons. For the measurements on solid solutions in anthracene it was found that at sufficiently high temperatures the conductivity became, in general, equal to that of pure anthracene, but at lower temperatures followed a similar exponential law with different values ϵ_1 and σ_1 of

the parameters. Figure 2 shows the graphs of log σ against $1/T$ for pure anthracene, and for a solution of naphthacene in anthracene of concentration 10^{-3}. The curves illustrate the main features found for pure hydrocarbons and for solid solutions respectively. Each pure material gave a single straight line with good reproducibility. The solid solutions of naphthacene and pentacene had conductivities which were equal to that of pure anthracene, within the experimental error, above a critical temperature which was characteristic of the solute and its concentration. Below this temperature the slope changed abruptly, giving values of ϵ_1 less than ϵ_0 for pure anthracene. The value of ϵ_1 for a given solute was independent of concentration. The plots do not fit a sum of two exponentials at all well, the break between the two branches being too sharp.

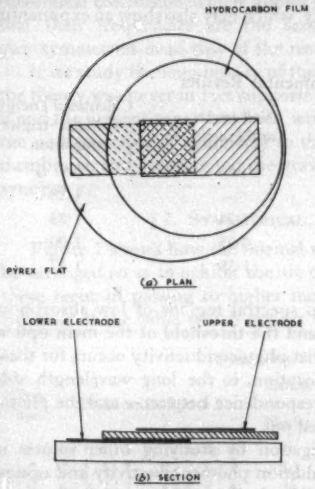


Figure 1. Diagram of the specimen geometry.

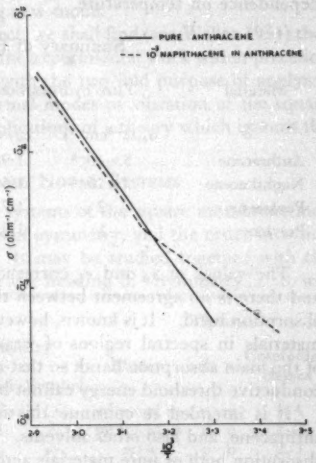


Figure 2.

In the low temperature branches for naphthacene and pentacene the higher concentration showed the greater conductivity. In the case of perylene, however, the conductivity of the 10^{-3} solution was less than that of pure anthracene throughout the accessible temperature range, whereas the conductivity of the 10^{-4} solution was always greater. The values of ϵ_1 were the same for both concentrations. This anomaly, which was reproducible, has not as yet received an explanation.

The table shows the experimental results so far obtained. The first two columns give the values of σ_0 and ϵ_0 for the pure hydrocarbons. The values for ϵ_0 are probably accurate to within $\pm 5\%$, but the values of σ_0 are less accurate because they depend on measurement of the thickness of the specimens, which could not be performed with high precision. In the third column are shown the values of ϵ_1 found from the low temperature branch of the curves obtained for

solid solutions in anthracene. Finally the fourth column shows, for comparison, the values of the singlet–triplet transition energy calculated by Hall (1952). These relate to the lowest excited states of the free molecule, and it would not, at first sight, seem reasonable to expect agreement between these energies and the thermal activation energy for the conduction process in the solid. At the same time there is certainly some correlation between the lowest excited states of the free molecules and the energy required to remove an electron from one molecule and place it on an unexcited molecule at a distance. The agreement between the figures in column four and the values of ϵ_1 for naphthacene, pentacene and perylene and ϵ_0 for anthracene is closer than would be expected on this basis, though it is significant that the values of ϵ_0 for the other pure hydrocarbons do not agree with the calculated singlet–triplet energies. We are not at this stage in a position to consider the effect of mobility, which may also show an exponential dependence on temperature.

Summary of Experimental Results

Material	Pure hydrocarbons		Solutes in anthracene	Calculated energy of singlet–triplet transition
	$\sigma_0(\Omega^{-1}cm^{-1})$	ϵ_0(ev)	ϵ^1(ev)	(ev)
Anthracene	5×10^{-2}	1·92	—	1·88
Naphthacene	2×10^{-1}	1·64	1·24	1·33
Pentacene	2	1·72	0·95	1·00
Perylene	1	1·93	1·58	1·57

The values of ϵ_0 and ϵ_1 correspond to spectral regions of low absorption, and there is no agreement between them and the threshold of the main optical absorption band. It is known, however, that photoconductivity occurs for these materials in spectral regions of weak absorption, to the long wavelength side of the main absorption band, so that a correspondence between ϵ and the photoconductive threshold energy cannot be ruled out.

It is intended to continue this investigation by studying other solutes in anthracene, and also other solvents. In addition photoconductivity and optical absorption both of pure materials and of solid solutions will be examined.

We are indebted to Professor J. W. Cook for the sample of perylene, and to Dr. E. Clar for the pentacene. Acknowledgment is also due to the Admiralty for permission to publish this work.

References

Bowen, E. J., Mikiewicz, E., and Smith, F. W., 1944, *Proc. Phys. Soc.* A, **62**, 26.
Hall, G. G., 1952, *Proc. Roy. Soc.* A, **213**, 113.
Inokuchi, H., 1954, *Bull. Chem. Soc. Japan*, **27**, 22.
Mette, H., and Pick, H., 1953, *Z. Phys.*, **134**, 566.
Simpson, O., and Northrop, D. C., 1954, *Physica, Proceedings of Amsterdam Conference, July* 1954, in the press.

Symmetry of Vibrating Square Membrane

By MARY D. WALLER
Private address: 5 Gloucester Gate, London N.W.1

MS. received 28th September 1954

§ 1. Introduction

It is now certain (see Waller 1952) that the vibrations of free square plates always conform with the symmetry of the surface.

But this experimental conclusion conflicts with the long established theoretical conclusion which allows two normal modes of vibration, provided that their frequencies are the same, to combine regardless of either their own symmetries or of that of the resulting new mode.

If we study the long history of the subject, we shall find (see Waller 1954) that the theory was never in fact supported by the experimental work which preceded it and the present paper has been written with the two-fold purpose of applying the new principle of symmetry to the normal modes of vibration of the square membrane, and pointing out the grave implications of a theory which ignores the symmetry.

§ 2. Symmetrical Normal Nodal Systems

Figure 1 shows how the normal nodal systems of the square membrane may be arranged so as to exhibit the six classes of symmetry, and the order in which these recur in passing to higher modes. It may be studied together with the columns of the table which appear under the heading of 'Symmetry'. It will

Class No.	Position in figure 1	Symmetry: Lines of symmetry: Nodal	Symmetry: Lines of symmetry: Anti-Nodal	Symmetry: Symbol	Theory: Mode	Theory: m	Theory: n	Combining classes
1	1 | 1, 5 | 1. . , 1 | 3, 5 | 3. .	0	2d, 2s	○	$m\|\ n+n\|\ m$	o	o	1, 4
2	2 | 1, 4 | 1. . , 2 | 3, 4 | 3. .	1d	1d	⁄	$m\|\ n-n\|\ m$	e	o	2, 3
3	1 | 2, 3 | 2. . , 1 | 4, 3 | 4. .	1s	1s	—	$m\|\ n$	e	o	2, 3
4	3 | 1, 7 | 1. . , 3 | 3, 7 | 3. .	2d	2s	×	$m\|\ n-n\|\ m$	o	o	1, 4
5	2 | 2, 6 | 2. . , 2 | 4, 6 | 4. .	2s	2d	+	$m\|\ n+n\|\ m$	e	e	5, 6
6	4 | 2, 8 | 2. . , 4 | 4, 8 | 4. .	2d, 2s	0	✳	$m\|\ n-n\|\ m$	e	e	5, 6

d, diagonal; s, median; e, even integer; o, odd integer.

be noted that the displacements of the vibrating surface are always 'balanced' about one or more of the lines of symmetry. When such a line is nodal, as for one diagonal in 2¦1 and one median in 1|2, the displacements at mirror-image positions on either side are equal and out of phase as indicated by the + and − signs. Antinodal lines of symmetry are defined as lines for which the displacements at mirror-image positions are equal and in phase. Both nodal and antinodal symmetry lines (which occur equally often) may be crossed by nodal lines. The six classes of vibration pattern recur in every $m=4$ columns and in every two rows, and the diagram, which may be continued indefinitely both downwards and to the right, includes all the normal nodal systems which conform with the symmetry of the square surface.

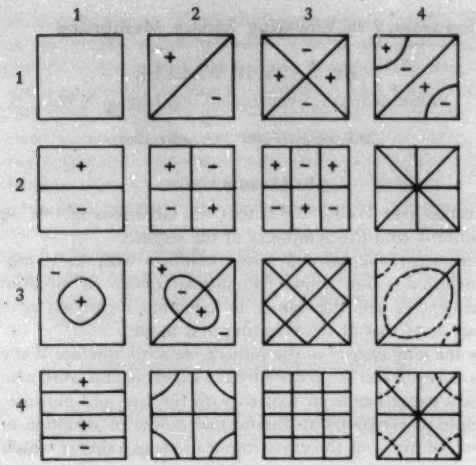


Figure 1. The six classes of symmetry recur in every two rows and four columns.

§ 3. Theory

The theory of the vibrating square membrane is given in standard works, e.g. Rayleigh's *Theory of Sound* (1894).

The well-known equation of motion states that

$$\frac{d^2z}{dt^2} = \frac{T}{\rho}\left(\frac{d^2z}{dx^2} + \frac{d^2z}{dy^2}\right) \qquad \ldots\ldots(1)$$

in which T is the tension and ρ the surface density.

The equation is linear and in the development of the theory it has been assumed that the principle of superposition can be applied to the standing waves of limited vibrating surfaces as well as to the progressive waves of an unlimited surface.

The nodal systems of a square membrane of side a are then given by

$$z = A \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{a} \pm B \sin\frac{n\pi x}{a} \sin\frac{m\pi y}{a} = 0 \qquad \ldots\ldots(2)$$

and the vibration frequencies by

$$\frac{1}{2a}\left\{\frac{T}{\rho}(m^2+n^2)\right\}^{1/2} \qquad \ldots\ldots(3)$$

where m and n are positive integers and the amplitude constants A and B may have any relative values.

It follows from expressions (2) and (3) that for given values of m and n every 'plus'† mode has the same frequency as the corresponding 'minus'† mode and that these pairs of vibrations may combine in any proportions.

† In theoretical work, the + and − modes are sometimes designated as 'symmetrical' and 'unsymmetrical', respectively, which terms will not be confused with 'symmetrical' as it is used in the present paper.

§ 4. Discussion

4.1. *Normal Nodal Systems*

The nodal systems of figure 1 (except those with only one nodal median) are given by putting $A=B$ in equation (2).

The various cases which may occur are summarized in the table under the heading 'Theory'. Class symbols have been included in the fifth column in order to facilitate the comparison between the later and the earlier columns of the table.

(*a*) When the values of m and n are either both even or both odd, as in classes 1, 4, 5 and 6, the symmetry is four-fold. Each term of the equation represents a similar 'primitive' vibration which is superposed at right angles to the other either in or out of phase. It is important to note that in the case of the plate these two standing waves cannot exist separately.

(*b*) When m is even and n odd, and vice versa, the equation represents the nodal system with one nodal diagonal, class 2. It so happens that the primitive vibrations are now identical with the normal modes of class 3 which are represented by one term of equation (2) and include one nodal median. This circumstance is 'accidental' and, although analytically classes 2 and 3 can each be derived from the other, it is of the greatest importance to realize that they are normal and not combined modes of vibration.

4.2. *Combined Nodal Systems*

The additional modes of vibration which are theoretically possible are given by making the amplitude constants A and B in equation (2) unequal.

Consideration shows that, as summarized in the last column of the table, for any given values of m and n, classes 1 and 4, 2 and 3, 5 and 6, respectively, should combine in pairs.

The textbooks, e.g. Barton (1914, fig. 52 and table IX) or Lamb (1931, fig. 49 and 50), give two examples of this conclusion which concern the first member of classes 2 and 3, and of 1 and 4, respectively.

(*a*) The first example is illustrated in figure 2 where the interior doubly curved dotted line through the centre is given by $B \simeq 2A$. It is noteworthy that while such a curve can be seen on a plate which is in the shape of a parallelogram it cannot occur on the square surface. The curve, like the parallelogram, possesses two-fold axial, but no line, symmetry.

Figure 2. Theoretical combined mode; frequencies equal.

(*b*) In the second textbook example (see 3|1 and 1|3, figure 1) designs are included which are proper to elliptical, or to wide rectangular, or to narrow rectangular, surfaces respectively, but impossible on the square plate.

(*c*) Similarly 4|2 and 2|4 (figure 1), the first members of classes 5 and 6, should combine, but the books do not explain how a vibration with two nodal median lines of symmetry combines with one in which all four lines of symmetry are nodal.

From the recent literature (e.g. Coulson 1952) it is doubtful whether the full implications of the above theory have been realized. It implies that the only

nodal systems in figure 1 which do not combine with another to form additional designs are the few $m=n$ systems which occupy the $1|1,\ 5|1,\ 9|1\ldots$ spaces in the first row and the $2|2,\ 6|2,\ 10|2\ldots$ spaces in the second row. The lower rows do not contain any patterns which should not combine.

Moreover equation (3) implies that if the sum m^2+n^2 is the same for different values of m and n, the two modes should combine because their frequencies are the same. Lamb (1931) instances the $7|4\pm4|7$ and $8|1\pm1|8$ modes. Another case concerns the $m=n=5$ and $7|1\pm1|7$ modes.

4.3. *A Significant Experiment with a Square Plate*

The fact that the present theory of vibrating membranes finds a place in analogies with certain problems in wave mechanics theory adds to the importance of the above discussion, and it has therefore seemed worth while to refer to a significant experiment with a square plate which has been described more fully by Waller (1952):

Let a square plate be set into free vibration in the first mode which includes a nodal median and then touched very lightly at one of the corners. Then the sand pattern will change over to the mode with one nodal diagonal (which has the same frequency) more rapidly than the eye can follow. It is equally easy to pass from a nodal diagonal to a nodal median by touching the midpoint of one of the sides. If, however, an attempt be made to produce a combined nodal design by touching the edge at some other point, the vibration will be damped out immediately.

Thus it appears that the energy of the vibration must reside in a normal mode which conforms with the symmetry of the surface. The slightest suitable disturbance suffices to change from one standing wave to the other but the energy cannot be shared nor can the two modes combine.

(Few 'square' plates are sufficiently perfect for this demonstration, but much can still be learnt if the four corners are numbered and all edges bowed in turn at different positions.)

§ 5. Conclusions

Now that Robinson and Stephens have obtained such excellent photographic records of vibrating circular soap films (Stephens and Bate 1950, fig. 5.34) it may be hoped that some reliable experimental results with square films may also, before long, become available.

In the meantime, there is every reason to accept that all the possible modes of vibration of the square membrane must conform with the symmetry of the surface. Indeed, the shape of the membrane (which determines the exact boundary conditions at the edges and the corners) is the only factor which can set any limit to the number of allowed modes in any given frequency interval.

References

Barton, E. H., 1914, *Text book of Sound* (London : Macmillan), §§ 189–197.

Coulson, C. A., 1952, *Waves : a Mathematical account of the common types of wave motion*, 6th Edn (Edinburgh and London : Oliver and Boyd), chap. III.

Lamb, H., 1932, *Dynamic Theory of Sound*, 2nd Edn (London : Edward Arnold), chap. V.

Rayleigh, Lord, 1894, *Theory of Sound* I, 2nd Edn (London : Macmillan), chaps. IX, X.

Stephens, R. W. B., and Bate, A. E., 1950, *Wave Motion and Sound* (London : Edward Arnold), fig. 5.34.

Waller, M. D., 1952, *Proc. Roy. Soc.* A, **211**, 265; 1953, *Acustica*, **3**, 370; 1954, *Ibid.*, in the press.

Note on Surface Vibrations of a Circumscribed Liquid

By MARY D. WALLER
Private Address: 5 Gloucester Gate, London N.W.1

MS. received 28th September 1954

It would be expected that the general law of symmetry which governs the vibrations of free plates (see Waller 1952) would also hold for the surface vibrations of liquids in containers sufficiently small for stationary wave patterns to form as the result of reflections at the boundaries. The law states that the normal modes of vibration of a plate with m lines of symmetry may be divided into $2f$ classes where f denotes the number of factors of m.

Most experiments on surface ripples, for instance those of Vincent (1897, 1898), have been made with large vessels where the edge effects are unimportant. The brothers Weber however (1825) published drawings of standing mercury ripples in a square container which include several of the possible six classes of symmetry. Again, Békésy (1932), using water, and both elliptical and triangular containers, gives illustrations exhibiting two-fold and three-fold symmetry, respectively.

Some of the examples of the law, with suggested shapes for the containers, are given in the table.

Shape of container	m line symmetry	Factors of m	f	$c=2f$
Rectangle, Ellipse	2	1, 2	2	4
Regular Triangle	3	1,3	2	4
Square	4	1, 2, 4	3	6
Hexagon	6	1, 2, 3, 6	4	8
Octagon	8	1, 2, 4, 8	4	8
Circle	∞	∞	∞	∞

f=number of factors; c=number of classes.

The periphery being nodal, the results in small square vessels might be compared with figure 1 of the preceding paper (Waller 1954) where the nodal systems of the six classes of the square membrane are represented. It might also be possible to obtain some comparable relative vibration frequencies.

An account and bibliography of the techniques which can be employed to produce and photograph liquid ripples is given by Brown (1936).

References

Békésy, G. V., 1932, *Z. Phys.*, **79**, 668.
Brown, R. C., 1936, *Proc. Phys. Soc.*, **48**, 312.
Vincent, J. H., 1897, *Phil. Mag.*, **43**, 411; *Ibid.*, 1898, **45**, 194.
Waller, M. D., 1952, *Proc. Roy. Soc.* A, **211**, 265; 1954, *Proc. Phys. Soc.* B, **67**, 895.
Weber, E. H., and Weber, W., 1825, *Wellenlehre* (Leipzig: Gerhard Fleischer).

REVIEWS OF BOOKS

Les Aberrations de l'œil—leur rôle dans l'accommodation, by A. IVANOV. Pp. 85. (Paris: Éditions de la Revue d'Optique Théorique et Instrumentale, 1952.) 560 fr.

This little book is written in three parts, dealing respectively with the theoretical eye and its aberrations, the experimental study of these aberrations, and their role in accommodation. The measurements described in the second part have already been published in a well-known series of papers by Professor Ivanov, and they would seem to constitute the most reliable results yet obtained. For the range $0{\cdot}45\,\mu$–$0{\cdot}70\,\mu$ the measured chromatic aberration is about 1·6 dioptres, and this agrees with the theoretical value to within a few per cent. The spherical aberration shows a marked variation with the state of accommodation. With 3 dioptres of accommodation the eye is over-corrected by the same amount that it is under-corrected when unaccommodated. Moreover, in all states, the focus is practically constant for rays at distances greater than about 0·5–0·7 mm from the axis. It is thus, in a sense, the centre of the pupil which is aberrated; the magnitude of this aberration varying from just under $+1{\cdot}0$ dioptre to about $-1{\cdot}0$ dioptre as the eye changes its focus from infinity down to 3 dioptres. Professor Ivanov attributes the spherical aberration of the eye to the non-uniform refractive index of the crystalline lens; and its variation to a redistribution of this non-uniformity arising from the change of shape of the crystalline lens in accommodating. It would seem desirable to study the influence of asphericities of the refracting surfaces on the spherical aberration before accepting this theory as proved. A simple calculation suffices to show that asphericities of the order of $50\,\lambda$ could produce a variation of spherical aberration equal to that found by measurement.

The final section of this interesting work shows how the aberrations of the eye assist in accommodating vision for near and far objects.

H. H. HOPKINS.

A Textbook of Physics, Vol. 3, *Sound*, by R. C. BROWN. Pp. iv+142. (London: Longmans, Green & Co., 1954.) 10*s*. 6*d*.

At the Intermediate or Advanced level, sound is often one of the neglected corners of the physics syllabus. Yet no branch of the subject is more rewarding to the teacher who looks ahead and seizes the opportunity to lay a thorough foundation of the physics of waves. This is the outlook of Dr. Brown's book, which forms the third part of his *Textbook of Physics*. The first three chapters give a clearly written account of the propagation and behaviour of waves, and are followed by an excellent chapter on forced vibrations and resonance. For the rest, the usual work on strings and pipes and measurements is adequately covered, with emphasis on principles rather than on details. The whole is written with a real understanding of the needs and difficulties of the average student, for whom it can be thoroughly recommended.

G. R. NOAKES.

Progress in Metal Physics 5, edited by B. CHALMERS and R. KING. Pp. vii + 324. (London : Pergamon Press, 1954.) 60*s*.

The new volume in the *Progress in Metal Physics* series contains articles on Fracture (N. J. Petch), the Plastic Deformation of Single Crystals (R. Maddin and N. K. Chen), the Structure of Liquid Metals (B. R. T. Frost), Precipitation (H. K. Hardy and T. J. Heal) and Solidification (Ursula M. Martius). Such articles as these, compiled from material drawn from a great mass of published work, are difficult to put together and require much laborious organization. A tribute should be paid to those who take on such tasks, for metal physics would be all the poorer to-day without the results of their work as embodied in the five volumes of this series.

The articles in the present volume are fairly formidable even by previous standards. The report on precipitation runs to 137 pages with 350 references. Its value to many will be great indeed, but it hardly constitutes the kind of review the Editors have in mind in their Foreword. Can someone who is not active in the field read this review through and obtain a fair appraisal of the present situation? It is a complex account of theory, technique and experiment with some of the characteristics of a catalogue, and whilst it is to be admired for its comprehensiveness, the subject is surely altogether too large for a progress article. In other words, it becomes simply a very good reference source. Now an article in such a book as this should, in the first place, be readable. It should be illuminated by a point of view, and it should be most rigorously pruned of material not directly relevant to the exposition: ample references will take care of the rest. One does not want a sequence of abstracted summaries, but cohesion and even conjecture; in fact, one wants to be aware of the progress and aware of the direction in which this progress is heading, the whole being held together by some central framework. Certain articles have achieved this in the past, and they have all been both readable and worthy works of reference: Cottrell on Dislocations (Vols. 1 and 4), Nowick on Internal Friction (Vol. 4), and the Polygonization articles (Vol. 2) come to mind, and there are others.

In the present volume, the fracture article fills a vacancy in this progress series. It is neatly compiled and very welcome. The consideration of the geometry of crystal deformation is usefully confined to crystallographic effects and particularly treats problems which have only recently been taken up as important—the heterogeneity of the glide process and composite slip for instance. The article gains by this limitation of viewpoint, as, in a way, does the article on solidification, which is severely confined to a consideration of some of the variables determining the mode of solidification. It does seem, however, to be quite inconclusive, and surely the whole question of nucleation and growth has received by now a disproportionate representation in this series, with the rate-process review (Vol. 2), the recrystallization and grain growth review (Vol. 3), and the nucleation review (Vol. 4), while other important subjects, such as fatigue, find no place among any of the thirty-four articles published so far. The liquid metals review introduces a new subject to the series, and reveals the inadequacy (indeed the naïvety) of the theoretical approaches to date. It would be interesting to see what a theoretical physicist, working on the subject, could have produced as a progress article of this nature. Fortunately, the emphasis is here on the experimental results, and as

(for some reason) the subject rarely achieves a comprehensive review, the present article is all the more useful on that account.

To sum up; a valuable volume, more prosaic than its predecessors in its general style, but a worthy addition to the reference value of this series.

A. J. KENNEDY.

Statistical Astronomy, by R. J. TRUMPLER and H. F. WEAVER. Pp. xx+644. (Berkeley and Los Angeles: University of California Press; Cambridge: University Press, 1953.) 56*s*. 6*d*.

In the development of stellar statistics, particular problems were often solved by methods developed for the purpose, and without reference to general statistical theory. This book attempts to relate the special problems and solutions of stellar statistics to the basic principle of modern statistical theory. The authors state in their preface that their object is to introduce the student to principal statistical problems in Astronomy, to their mathematical formulation, and to methods and techniques of their solution. This programme is carried out in six Parts, of which Part I, Elements of Statistical Theory, occupies one third of the volume. In this Part, statistical theory is developed *ab initio* in a manner suited to the specific needs of the statistical astronomer. In Parts II to VI, the theory is applied to the study of the structure, constitution and dynamics of our galactic star system. Part II, Statistical Description of the Galactic System, is a short general analysis of the scope of the applications of the statistical theory. This is followed in Part III by a discussion of Stellar Motions in the vicinity of the Sun, a field in which statistics found early and successful application. Part IV deals with the Luminosity-Spectral Type Distribution in which the frequency function according to the absolute magnitude and spectral type for stars in the vicinity of the sun is considered. Part V, Space Distribution of Stars, is an authoritative discussion in which the earlier work on star counts is reviewed, and a clear account is given of the more recent methods in which the interstellar absorption of light and the space density of the stars are determined from magnitudes, spectral types and colour indices. Part VI contains a useful discussion of galactic rotation which, however, allows very little scope for the application of the statistical methods elaborated in Part I.

Much of the contents of Part I would normally be covered in an undergraduate course in statistics, but the authors have chosen to provide for the needs of students who have managed to evade the discipline of mathematical statistics. In so doing, they have considerably lightened the student's task by teaching basic statistical theory and the applications to astronomy in a consistent notation. In propounding the theory, considerable attention is paid to numerical examples. The examples are worked out in full numerical detail and are of greater intrinsic interest for a student preparing for research in statistical astronomy than, for example, a well-known problem in which x denotes the length of the carapace of the common shrimp and y denotes the length of the postspinous portion of the carapace.

In Parts II to VI, the authors have chosen what is clearly a logical order of presentation of the topics within each Part. The historical trend can, however, be readily traced in the Notes and Bibliography with which each Part concludes. The usual apology is made in the Preface that the references are not a complete bibliography. But there can be no excuse for the absence in Part VI of any

mention of G. L. Camm to whom we are much indebted for the introduction of a powerful general method of analysing radial velocities for the effect of galactic rotation. The method is, in fact, very briefly described on page 556 without reference either to Camm or his work on the planetary nebulae and cepheid variables*. As an example of an error in the other direction, mention may be made of the reference on page 613 to some work on the force perpendicular to the galactic plane which is stated to differ considerably from Oort's results. The bibliography, however, shows that the work in question is contained in an unpublished thesis of the University of California written some nine years ago.

The level of exposition throughout the book is fairly uniformly high, and the clarity of the presentation is aided by thoughtful and well-drawn diagrams. The book is clearly a very useful addition to existing literature on the subject.

It is pleasing to see that so large and valuable a book has been produced at so reasonable a price. This has been made possible by subsidies from the University of California and the National Academy of Sciences.

D. W. N. STIBBS.

Wool — its Chemistry and Physics, by P. ALEXANDER and R. F. HUDSON. Pp. ix + 404. (London: Chapman & Hall, 1954.) 45*s*.

As the authors point out in the preface to this book many physical and physico-chemical theories and processes were first studied by scientists working on wool before being extended to other systems. Much of this work is reported in specialized journals which are not normally consulted by scientists working outside the textile field. This book fulfils a valuable function in bringing this work into a single volume. The value of the book in this respect is enhanced by a clear index and bibliographies at the end of each chapter.

It must, however, be stated at the outset that the title of the book is a little misleading as the author's main concern is with the physical and organic chemistry of the wool fibre. The general physical properties of wool are only mentioned in so far as they have been used in elucidating the chemistry or structure of the wool fibre. Reference to physical properties is confined largely to chapters on surface and mechanical properties, swelling and absorption. Even here there are notable omissions: thus in the chapter on surface properties no mention is made of wetting. There is also no reference to the behaviour of wool in torsion or to its visco-elastic properties. The important thermal properties of wool are completely overlooked apart from the thermodynamics of water absorption.

There is one serious error in this part of the book which should be corrected. It is stated (p. 23) that Medley found that the production of frictional electricity is independent of rubbing speed and attributed it to corona discharge.

Approximately half the book deals with the physical chemistry of the wool fibre, more especially with absorption phenomena. The various theories of water absorption are considered in some detail in the chapter on absorption and swelling whilst later chapters deal with the absorption of more complex materials such as electrolytes and dyestuffs. Particular attention is given to the Donnan and Gilbert and Rideal theories of electrolyte absorption.

The diffusion of different absorbates into the wool fibre is complicated by a diffusion coefficient which is usually dependent on the concentration of the

* *Mon. Not. R. Astr. Soc.*, 1938, **99**, 267; **104**, 163.

diffusing substance. This difficulty has led to developments in the mathematical theories of diffusion and the use of new techniques in their solution. Considerable space is devoted to these problems.

The remaining third of the book is devoted to the organic and stereo-chemistry of the wool fibre and includes an account of the later theories of the α-fold in keratin. Infra-red and x-ray techniques are also discussed.

The authors have not hesitated to express their own views on many controversial topics, but as opposing views are also discussed, this should make the book more stimulating to other workers.

Inevitably books of this nature dealing with rapidly expanding subjects are somewhat dated by the time they are published and no reference is made to work done within the last two years.

Altogether the book provides a useful survey on the physical and general chemistry of wool which should prove of value not only to research workers in many fields of physical chemistry but also to students.

The book is well produced with excellent illustrations.

F. L. WARBURTON

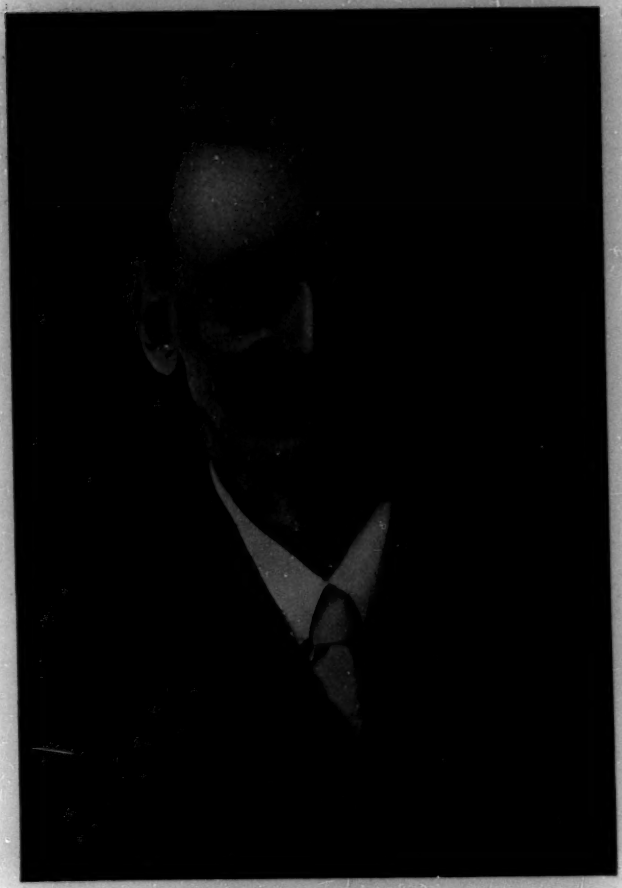
BOOK NOTICES

Inter-relations between the Terrestrial Atmosphere and Cosmic Radiation, by P. K. Bhattacharya. Pp. iv + 50. (Calcutta: Indian Physical Society, 1952.)

Proceedings of the Second International Congress on Rheology, held at Oxford on 26th–31st July, 1953, edited by V. G. W. Harrison. Pp. ix + 451. (London: Butterworth; New York: Academic Press, 1954.) 60*s*.

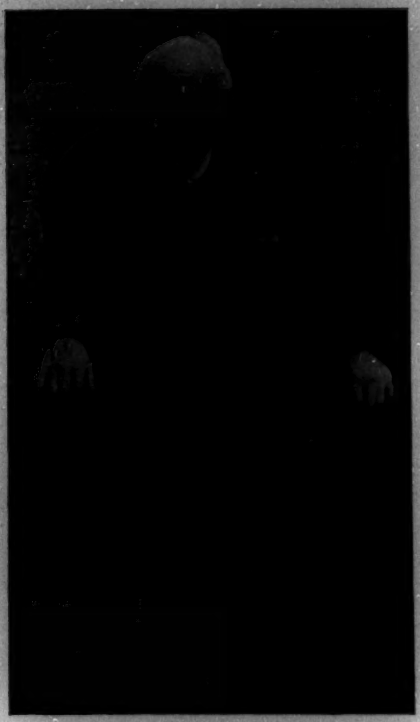
Essai sur la convection naturelle, by P. Vernotte. Pp. xi + 58. (Paris: Publications Scientifiques et Téchniques du Ministère de l'Air, No. 288, 1953.) 600 fr.

Contribution a l'étude de la transmission infrarouge de la basse atmosphere, by Pierre Barchewitz, G. Amat and C. Rossetti. Pp. 90. (Paris: Publications Scientifiques et Téchniques du Ministère de l'Air, Bulletins des Services Téchniques, No. 116, 1954.) 800 fr.

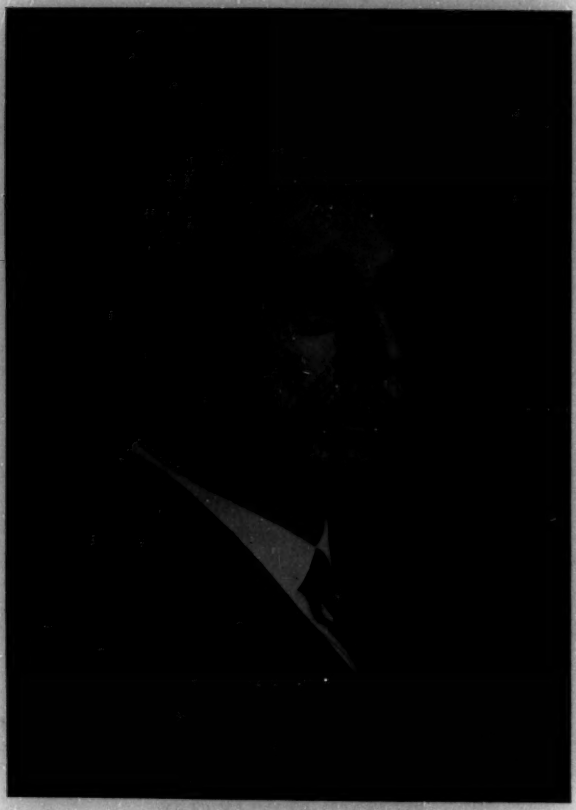


ALBERT CAMPBELL.

F. C. CHALKLIN.



REGINALD STANLEY CLAY.



ROBERT STEWART WHIPPLE.

OBITUARY NOTICES

HERBERT STANLEY ALLEN

Herbert Stanley Allen, the fifth son of Rev. Richard Allen, a Methodist minister, was born at Bodmin, Cornwall on 29th December 1873. During his early education at Kingswood School, Bath, he already showed great promise of a distinguished career. He was a Senior Prefect of the School and gained the distinction of taking the first place for all England in the London Matriculation Examination and in the Senior Cambridge Local Examination.

At Trinity College, Cambridge, which he entered in 1893, he was awarded a First Class in the second part of the Natural Sciences Tripos. He then went to Aberystwyth where he spent a short time in a temporary post as assistant lecturer before returning to Cambridge as a research worker under Professor J. J. Thomson. In 1900 he went to Renfrew to take charge of Lord Blythswood's Physical Laboratory. In 1905 he was appointed to the staff of the Physics Department in King's College, London, where in 1909 he was awarded the degree of D.Sc. for his work on the photoelectric effect.

In 1907 he married Miss Jessie Macturk whom he had met in Renfrew. They settled in New Malden, Surrey, where both their children were born and where they stayed until he was appointed lecturer in Professor Barkla's Department in Edinburgh University in 1920. There he became a Reader. After spending three years in Edinburgh he was appointed to the Chair of Natural Philosophy in The United College of St. Salvator and St. Leonard in the University of St. Andrews.

H. Stanley Allen was, for nearly fifty years, a member of the Physical Society. He took an active interest in the Society and between 1916 and 1921 he was a Member of Council and Honorary Papers Secretary. The Society received with deep regret the news of his death on 27th April 1954.

He was elected to the Fellowship of the Royal Society of Edinburgh when he took up his post in that city. He was awarded the Makdougall-Brisbane Medal and served on the Council. Election to the Fellowship of the Royal Society followed in 1930. This honour he shared with his brother Edgar Johnston Allen who was for many years director of the Marine Biological Laboratory at Plymouth. When, in 1944, he retired from the Chair of Natural Philosophy, the University of St. Andrews elected him Professor Emeritus and conferred on him the honorary degree of LL.D., fitting tributes to his devoted service to the University and to the cause of science.

His enthusiasm for research led him into many branches of physics among which may be mentioned photoelectricity, the Zeeman effect, spectroscopy, x-rays and radioactivity. In St. Andrews he was chiefly interested in the quantum theory and its applications to spectroscopy, especially to the band spectrum of hydrogen. Apart from publishing many research papers he was the author or joint author of a number of books. The first of these, published in 1913, was his *Photoelectricity*, based largely on lectures delivered in London and including an account of his own work on photoelectric fatigue.

The representation of physical phenomena merely in terms of a mathematical expression was, to Professor Allen, unsatisfying, and in his writing he always

tried to adopt the middle course between a purely mathematical treatment on the one hand and a purely descriptive treatment on the other. Whenever possible he tried to picture a physical model in terms of which the nature of the processes involved might be envisaged. This ' middle course ' he mentions in the preface to *The Quantum and its Interpretation* (1928), a book which did not profess to be a treatise on the quantum but rather to be " an attempt to deal with the baffling problem of the nature of the quantum ". This book was in the press just when the results of the earliest experiments on electron diffraction were being published. At that time much thought was directed to the dilemma of particles versus waves, a subject in which Allen was intensely interested and one which forms the theme of *Electrons and Waves* which appeared in 1932. In this book which is in a more descriptive vein he expressed his faith in an ultimate explanation of this dilemma. " Truth ", he says, " is great, and will prevail ".

Allen and Moore's *Text Book of Practical Physics*, first published in 1916, earned a well-merited popularity and by 1948, when the third edition appeared, it had passed through frequent reprinting. In *A Text Book of Heat* (1939), written in collaboration with his nephew, R. S. Maxwell, he adopted the historical method of presentation which is not so common as it deserves to be in scientific text books today. He had a gift for lucid exposition which was no less apparent in his lectures than in his writing; he spared no effort to make his meaning clear and, above all, to be accurate.

In his youth he was fond of cycling and walking; he and Mrs. Allen spent many happy holidays on the Continent. In later years he enjoyed motoring with his family in the Highlands of Scotland although he himself never mastered the art of driving. Professor Allen was naturally shy and retiring. He was always ready to help others with encouragement and advice; his personal qualities and the inspiration of his teaching won for him the respect and affection of his students and of his more immediate associates. It was only those who knew him most intimately who could appreciate the kindness which lay behind his quiet dignity of manner.

There are many who mourn the death of one who so unsparingly devoted himself to his life's work. Professor Allen died in the Black Isle, Ross-shire, at the home of his daughter Mrs. G. S. M. Walker with whom he spent the last year of his retirement.

D. JACK.

ALFRED BLACKIE

Alfred Blackie was born in 1882 at Crowborough, Sussex. He was educated at Tonbridge School, and Peterhouse, Cambridge. He obtained a First Class Honours Natural Sciences Tripos in 1903, and was awarded the Wiltshire prize for Mineralogy and Geology. Subsequently in 1905 he took the Mechanical Sciences Tripos.

His first post was on the teaching staff of the Royal Naval College, Dartmouth, an association which he very much treasured. In 1907 he joined the Physics Division of the National Physical Laboratory. His special knowledge of crystallography was of value in an investigation into the spontaneous changes of crystalline form that silica undergoes on prolonged heating at high temperature.

While at the N.P.L. he also devised a means, subsequently used in the Indian survey, of determining the corrections for the variation with height of the refractive index of the atmosphere.

As a territorial officer in the Artists Rifles he was mobilized on the outbreak of war, and from 1914 to 1919 served with the Royal Engineers on Coast Defence in Southern Ireland and on the Channel coast, with the rank of Captain.

On demobilization in 1919 he joined the newly founded Fuel Research Station of D.S.I.R. at Greenwich under its first Director—Sir George Beilby. In charge of the Physics Department he was responsible for developing a number of ingenious instruments, and methods of measurement of radiation and temperature, which proved of great value in the work of the Station. In research work on domestic heating he undertook fundamental work on the relationship of the optical density of smoke and its mass concentration, and on methods of reducing smoke. His advice was sought during World War II on smoke production for camouflage purposes. Before his retirement in 1947 he was responsible for designing a unique calorimeter building at the Fuel Research Station, containing 4 rooms, 12 ft. by 12 ft. by 9 ft., each a calorimeter equipped to measure completely and accurately the heat emission of a domestic heating appliance.

His work was marked by a care and accuracy which is the hall mark of the true physicist. His wide knowledge in the field of natural science was a help to all his colleagues. He was a Life Fellow of the Physical Society, and a Founder Fellow and Life Member of the Institute of Physics.

His interests were wide. He enjoyed foreign travel; he had a great love of nature, and specialized in Lepidoptera. He was a charming companion and friend, always willing to help those in need. Although not robust in health, he insisted on joining the Home Guard, and served as a Captain in the 23rd County of London Battalion. He died after a short illness in his 72nd year.

A. C. M.

ALBERT CAMPBELL

Albert Campbell, one of the oldest Fellows of the Physical Society, a Vice-President from 1910 to 1912, for many years a frequent contributor to the *Proceedings*, and the Duddell Medallist for 1925, died at Cambridge on 6th February at the age of ninety-one. He will probably be best remembered for his pioneer work on electrical measurements at the National Physical Laboratory, in particular his primary standard of mutual inductance and his alternating current bridges, but those who met him will long retain lively recollections of a highly original personality with qualities that aroused the warmest feelings on all sides.

He was born at Ballynagard House near Londonderry, the third son of Thomas Callender Campbell, J.P., who seems to have been a man of some enterprise with varied interests that must have done much to stimulate in his son the taste for practical scientific work. At Ballynagard he ran a mill for scutching the flax grown in the neighbourhood, grew gooseberries in quantity to supply a Glasgow jam-maker, and later, by way of producing a fertilizer as a substitute for imported Peruvian guano, set up the necessary plant, including

a mill for grinding bones, and a complete installation for synthesizing sulphuric acid by the lead chamber process. The " manure works ", wrote Albert, " served the country for many years ", but the family had the option of giving up the lease of the house and land every seven years, and " in 1898 we used this option, dismantling the manure works and taking down the tall chimney. We got a site nearer town, where we built the house Ballynatrua partly with material salvaged from the manure works ". Albert's interests were to run on different lines, but it is possible to discern something of the same pattern in them.

When he left the N.P.L. in 1918 he went back to Ballynatrua and it was there that he made his last and best absolute determination of the ohm, and from there he published it in what must be one of the shortest papers ever printed in the *Proceedings of the Royal Society*. In a private letter (Nov. 28, 1924) he gave a few more details. " The result came out $1{\cdot}0005_4$, but the probable error is at least $\pm 0{\cdot}0001$, for my time measurements are inadequate. I used a 100 c/s fork which I had tested before leaving Teddington, but I checked it with a phonic wheel and a 100-year-old 8-day clock which belonged to my great-grandmother."

But to return to his beginnings: he received his early education at Derry Academy, Queen's College, Belfast, and Edinburgh University. The link with Edinburgh no doubt came from his mother's family, who were Callenders of Leith. His first paper was a note on the Peltier effect, presented to the Royal Society of Edinburgh in 1882, presumably while he was still a student under Professor Tait at Edinburgh University. In the following year he proceeded to Cambridge; he was admitted at Corpus Christi College in October 1883, became a Foundation Scholar in 1884, headed the Senior Optimes in Part I of the Mathematical Tripos, 1886, and got a Second in the Natural Sciences Tripos, 1887. His energy at this time had clearly not been all devoted to work for the Tripos, for his second paper on the Peltier effect was published in July 1887, and he mentions that the experiments were carried out in Professor Tait's laboratory and later continued in his own laboratory at Ballynagard. However, he had not yet discovered the line of work that he was to make his own. He became a science master at various schools, took a short course in electrical engineering at the Central Technical College, South Kensington, and then in 1893, at the instigation of Alexander Russell, he joined the staff of Faraday House. Here he found himself in his element; he took part in the teaching work and supervised the testing and thus acquired an insight into the scientific problems of the rapidly growing electrical industry. Original work on instruments and measurements and on the electrical and magnetic properties of materials immediately followed, and by the time he left Faraday House he had published about a dozen papers, no less than five being read before the Physical Society between 1896 and 1901. However his best-known work was to follow.

In 1901 Glazebrook entrusted him with the task of building up in the newly-founded National Physical Laboratory a section for the measurement of the basic quantities encountered in alternating-current practice—inductance, capacitance, frequency, etc. He started from zero and during the next eighteen years gradually developed instruments and methods of measurement that were recognized as amongst the best of their kind in the world. The original Campbell primary standard of mutual inductance still forms the ultimate reference standard for the country, and the standards employed in several other

national laboratories are based upon it. The Campbell inductometer or variable standard of mutual inductance was at one time a key-instrument in almost every laboratory and test-room concerned with a.c. measurements, and his vibration galvanometer, his constant-inductance rheostat, and later his a.c. potentiometer became also very widely used.

Most of his instruments were made with his own hands. He found great satisfaction in such work and even made his own galvanometers and resistance boxes, though he was not a skilled mechanic in the ordinary sense of the term. He brought fine craftsmanship to bear on the vital part of an instrument, but everything else would be of the most primitive construction, and to any user but their maker the auxiliary mechanical adjustments were often almost comically crude.

This was all of a piece with his general character, which was one of extreme simplicity. Looking back, one associates most of his colleagues with brilliant careers; of Campbell no such thought is possible; one sees only a single-minded devotion to an ideal, not perhaps formulated explicitly, but unmistakably including simplicity as an essential element. It appeared not only in his instrument design but also in his gardening and music, which seemed to rank as the counterparts of his scientific work, his devotion to truth as he saw it in the one being matched by that to beauty in the others. He had very strong feelings about things that to him did not ring true, and he expressed them with force and wit. They might be anything from a part of an instrument showing a waste of labour or material, a pointless bit of administrative routine, or " the hideous wobble of the professional singer " (vibrato). A characteristic remark in gentler vein occurs in some memories of his childhood. Of his father, " He was a constructive gardener, not like my Mother, who was mainly a destructive one. She loved *tidying up* ".

Such a personality could not fail to find much that was uncongenial in a large highly organized semi-industrial laboratory and as the National Physical Laboratory continued to expand there came a point where Campbell felt that he would be better outside (though to the last his interest in the N.P.L. was always that of one who belonged to it). He was unmarried; his wants were simple and he estimated that he could manage to live without his salary, so he went back to Ireland and lived with his mother at Ballynatrua. He set up there his own laboratory and continued his work. His determination of the ohm there has already been mentioned.

On his mother's death in 1925 he decided to move to Cambridge. He packed up his laboratory and library (in 150 boxes !) and continued to work for about a year in lodgings while he had a house built to meet his own requirements. The ' base line ' was a 2-metre scale distance that, following his N.P.L. practice, he required in the principal room for his galvanometers. A treadle lathe stood on one side and on the other a doorway led into the greenhouse where he raised his seedlings. In this setting he steadily followed his chosen course to the end.

It is not possible to give in a few words an adequate impression of his total achievement. Much of it was summarized in the articles he wrote in Ireland for Glazebrook's *Dictionary of Applied Physics*, Vol. 2. He was urged to expand these into a book but heavy literary work did not greatly appeal to him and it was a relief to him when he found in Dr. E. C. Childs a man with whom he could collaborate. It is a satisfaction to note that the book was published in 1935.

His closing years were uneventful. He was no longer to be met at scientific meetings but occasionally appreciative fellow workers from as far afield as the United States and Japan would seek him out, an informal scientific honour that one felt to be peculiarly appropriate.

L. HARTSHORN.

FRANCIS CECIL CHALKLIN

The tragic death of Frank Chalklin at the age of fifty-one came as a profound shock to his many friends. He was Professor of Physics at Canterbury University College, New Zealand, and had applied for the vacant chair of physics at University College, Hull. The authorities at Hull desired an interview and arranged for Chalklin to fly to England for this purpose. He duly embarked on the Quantas–B.O.A.C. Constellation air liner which crashed at Kallang airport on 13th March, 1954, there being no survivors among the thirty-one passengers. He had left his home in high spirits and the end was mercifully sudden.

Frank Chalklin was born on 15th November, 1902, at Hadlow, near Tonbridge, of Kentish stock on both sides of the family, and his affection for his native county was very strong. He went to Judd School, Tonbridge, which he left in 1919 with high distinctions. He then studied at King's College, London, and after graduating in 1923 carried out research in the Wheatstone Laboratory, then under the guidance of Sir Owen Richardson. The subject of his early investigations was the excitation levels of very soft X-rays, on which three papers by Richardson and Chalklin appeared in the *Proceedings of the Royal Society*. The experience which Chalklin obtained in these early days, both in experiment and theory, stood him in good stead during his whole research career, for very soft X-rays and allied subjects remained his great laboratory interest throughout his life. The method used in the King's College experiments was a photoelectric one, but Chalklin subsequently turned to vacuum spectroscopy, with which he had outstanding success. For his work under Richardson he was awarded the Ph.D. in 1926.

From 1927 to 1930 he was at the University of Sheffield as junior lecturer under Professor S. R. Milner, whence he came to London to take up the post of lecturer on the physics staff at University College. Here he speedily established his position both as a teacher and as an original worker. His unselfishness and integrity made him both loved and respected: nothing was too much trouble if he thought that duty demanded it. Quiet and unassuming, he had beneath the surface unshakable standards to which he remained true. He was thorough and dependable in everything that he did, with no tricks for seeking the popularity which he nevertheless obtained. At University College, in conjunction with his wife, an accomplished physicist, he developed his first vacuum spectrometer, making use of a narrow ruled grating at grazing incidence. With this apparatus outstanding results were obtained in the neighbourhood of 50 Å, a region then almost unexplored. He speedily became an acknowledged leader in this field, his discussion of his results showing a command of the theoretical aspect of his subject. In 1938 he published, with collaborators, in our *Proceedings* an account of a new vacuum spectrograph with a concave grating, designed for the spectral region 15 to 1000 Å. His work with this instrument was brought to an end by the war, which saw the destruction of the

University College Physics Laboratory. During this period at University College he was awarded the London D.Sc. and made Reader in Physics of the University.

The war years he spent at the University College of Wales, Aberystwyth, whither part of the physics department of University College had emigrated. Here he won golden opinions by his knowledge, tact and assiduity. The responsibility exercised a maturing influence. As soon as he returned to the desolation at London in 1944 he threw himself wholeheartedly into aiding with the work of restoration and speedily had his own research started again, as witness a considerable paper on intensity measurements in the very soft x-ray region which appeared in the *Proceedings of the Royal Society* in 1948. At this time Chalklin was already in New Zealand, having been appointed to his Canterbury chair in 1946.

In New Zealand he set to work with his accustomed energy and with a full measure of success. He started research there by building, with his research students, the necessary apparatus, including the Geiger counter spectrograph for soft x-rays described in our *Proceedings* in 1951. He worked with system, laying the foundations surely for a solid school. In 1951 he was elected Fellow of the Royal Society of New Zealand and he had made himself greatly loved. The very full attendance at the memorial service held in the Cathedral at Christchurch bore witness to the respect in which he was held as a man and as a scientist.

Frank Chalklin had a very happy home life. In 1928, while he was at Sheffield, he married Letitia (Letty) Pulley Davies, who, as already recorded, helped him as collaborator in some of his best work and in less learned ways no less. There are three children, the eldest, a boy, now reading history at Oxford, and two girls, the youngest born in New Zealand. He was a very human character, much concerned with cricket, and himself a player of lawn tennis and badminton. Farming, too, was in his blood and in New Zealand the Chalklins had a small farm, run by a manager but the object of his expert interest. This notice can fitly conclude with some words which he himself wrote, just before his death, on his conception of a professor's task. They indicate well the man. "The basic duty is, of course, to maintain a good standard of undergraduate teaching. And I also hold the orthodox view that the university is still the best place for performing fundamental pure research and indeed has the duty of conducting it. I conform also with the view that the teaching is likely to be uninspired in departments in which research is neglected. The professor would seem to have the duty of encouraging his staff in this direction and, in particular, of helping young staff members to get the habit of research early in their careers. Their research record enables them to compete for posts in other universities as well as to secure promotion in their own. And movement between universities is stimulating to the individual and to the department. Finally, I believe that it is a duty to make the department a happy place for staff and students and that the senior and research students should be encouraged to look upon their department as their home."

E. N. DA C. ANDRADE.

REGINALD STANLEY CLAY

The death, on 10th April 1954, of Dr. R. S. Clay at the age of 85 years, removes a much respected figure from scientific circles in London.

Reginald Stanley Clay was born at Portsmouth and was educated in the first instance at Tollington Park College. He took the degree of B.Sc. (London) in 1889, and afterwards went to St. John's College, Cambridge, where he read Mathematics, becoming 21st Wrangler in 1892. Two years' experience as Physics Master at Mill Hill School was followed by his appointment in 1897 as head of the Physics Department at Birkbeck College. He was awarded the degree of D.Sc. (London) in 1900, the title of his thesis being 'On the application of Maxwell's curves to three-colour work, with especial reference to the inks to be employed, and to the determination of suitable light filters'. However, he seemed to incline to the fields of teaching and administration, rather than to the exclusive pursuit of science, and in the same year he accepted the post of principal of Wandsworth Technical College. Only two years later he was appointed Principal of the Northern Polytechnic Institute.

His position now allowed him the opportunity to wield a great influence in the development of the many sides of educational work in London, by no means exclusively concerned with science and technology. For example he was an active member of the National Council for Domestic Studies; he was Chairman of the Governors of the North London Collegiate School for Girls; and he became a member of various societies with cultural interests. From 1907 to 1911 he acted as secretary of the Association of Technical Institutes. A lifetime of social service was the outward expression of his Christian faith. He was a member of a Congregational Church, and became a Mason in the Ophthalmos Lodge in 1929.

His major scientific interest lay in the field of Optics, and his book on *Practical Light*, published in 1911, was widely used and valued. He joined the Optical Society in its early days, and took a large share in its elevation (during the 1914–18 war) from a minor status into that of a leading scientific Society. He served on its Council, and was President from 1927 to 1929. His Presidential Address dealt with the History of the Stereoscope. He was a keen participant in the discussions (sometimes rather heated) by which the members accomplished a good deal of their mutual education in optical matters. But Clay's urbane manner and charming smile never failed him. His numerous duties left him little time for continuous research, and he published comparatively little in this period beyond occasional brief descriptions of apparatus, but he was almost the ideal member and critic. Many were the occasions when some ingenious piece of equipment made by his own hands was brought to a meeting for exhibition. He was uniformly kind in his encouragement of younger workers, many of whom had cause to be grateful for his hospitality and constructive hints.

He took much interest (as representing the Optical Society) in the formation of the Institute of Physics, being a Founder Fellow in 1918. He was elected to the Board in 1920, and served the Institute in various capacities during the remainder of his life, i.e. as Examiner and as Chairman of the Advisory Board

for the *Journal of Scientific Instruments*, to which he was a frequent contributor of short articles describing various tools and processes. He was also a member of the Council of the British Optical Association, and was its President in 1927; and he acted for many years as Examiner for this body.

He became a member of the Physical Society in 1925, and later of the Optical Group, in which he took a great interest. He was welcome and honoured at all social events in connection with these bodies.

His retirement from the Northern Polytechnic in 1931 gave him more leisure to pursue his favourite studies. Somewhat before this time, he had become acquainted with Thomas Henry Court, and had begun to collaborate with him in historical studies of optical instruments, especially the microscope. (He was already a member of the Royal Microscopical Society, of which he served as President in 1937.) The most noteworthy product of this friendship was the *History of the Microscope*, written in collaboration with Court, and published in 1932. Clay himself had for some time been collecting old instruments, and in the end brought together a very representative set in which the period from A.D. 1670 to 1850 was represented by three hundred and twenty-eight microscopes. It is not surprising that the study of these, together with Court's own magnificent collection, made the foundation for a most authoritative book. The Clay collection was acquired in 1943–44 for the Ashmolean Museum, Oxford.

In 1936 he joined the Board of the United Kingdom Optical Company Ltd., at Mill Hill, and was associated with this firm in various ways till his death. During the war of 1939–45 he had much to do with the Admiralty Research Laboratory at Teddington, acting as liaison officer between this and various Service depots.

One of his greatest pleasures in more recent years was his membership of the Royal Institution; he was a regular listener to the Friday evening discourses, and took the greatest interest in the demonstrations. His faculties seemed undimmed and his memory keen; his delight in science as great as ever. It was difficult to believe that he was approaching his 86th year. He suffered only a short illness.

He married Theodora Tilley, who died in 1952. They had a son and two daughters.

The passing of a man such as this marks an epoch in the development of the Physical Sciences. At its beginning, the keen amateur could hold his own and take an important part; at its end, the Scientific Society largely reflects an aspect of professional duty. Is there any room for the non-specialist? It seems certain that, without the spirit of such people, Science might prove to be one of the biggest of human disappointments.

L. C. MARTIN.

FRANK KENNETH GOWARD

Frank Kenneth Goward, whose untimely death occurred on 10th March 1954, occupied a post of Senior Principal Scientific Officer at the Atomic Energy Research Establishment, Harwell. Born on 30th August 1919, he received his early education at Queen Elizabeth's Grammar School, Wakefield, and from there went up to St. John's College, Cambridge, in 1938 on an open scholarship. After graduating in honours physics in 1940 he joined a group of physicists working on the radar chain stations in the early years of the war. With this experience behind him he joined the Telecommunications Research Establishment at Malvern towards the end of 1942. From then until the end of the war he was concerned with research and development of centimetre aerial systems and rapidly established a reputation as an authority in this field.

When the war finished there was intense interest in this country and in the United States in methods of particle acceleration using radio-frequency fields. Frank Goward was quick to realize that experimental verification of the new principles of synchrotron acceleration could most readily be obtained by conversion of an existing betatron, and in collaboration with D. E. Barnes was the first to accelerate electrons by this method to an energy of 8 MeV. Higher energy synchrotrons soon came into operation in other laboratories and Goward played a leading part in the design of Britain's largest electron accelerators at Glasgow and Oxford Universities, and in several 30 MeV machines built at the Atomic Energy Research Establishment. Using the latter machine Goward and his team made a number of notable contributions to fundamental nuclear research in their photodisintegration work using photographic plate techniques.

In 1951, Goward was chosen to represent Great Britain at an international conference held under the auspices of Unesco to discuss the formation of an international laboratory for Nuclear Research. Arising from this he was largely responsible for drawing together an international team of scientists to examine the problems of design of a 15 kMeV proton synchrotron. The activities of the group entered a new and exciting phase in mid-1952 after Goward and the group leader Dr. Dahl brought back from America information of the new principle of alternating gradient focusing. In 1953 Goward was given leave of absence from Harwell and moved to Geneva with his family to take up his duties as deputy group director of the proton synchrotron team for C.E.R.N. (the Central European Laboratory for Nuclear Research). Not many months afterwards he became seriously ill and his wife brought him back to this country for specialist treatment. His colleagues will remember Frank Goward for his irrepressible sense of fun and good humour.

W. WALKINSHAW.

FRANK LLOYD HOPWOOD

The death on 2nd May 1954 at the age of 70 of Professor F. L. Hopwood, one of the oldest members of the Physical Society and for many years a member of its Council, is a great loss to medical physics of which he was a leading authority.

Born in 1884 in Cheshire, Frank Lloyd Hopwood received his education at Hawarden Grammar School, the University College of North Wales, Bangor, and the Royal College of Science, and finally at the University College London, where he received his D.Sc. degree in 1919. He joined the staff of St. Bartholomew's Hospital in 1906 and was teaching physics to the students at the Medical College there. In 1920 he became Physicist to the Hospital, and in 1924 was appointed to the Chair of Physics at the Medical College of St. Bartholomew's Hospital, a post he held until his retirement in 1949, when the University of London conferred on him the title of Professor Emeritus.

During the First World War Hopwood worked with Sir William Bragg on submarine detection problems. Ever since then he maintained an interest in ultrasonics, and he pursued research in this subject and published several papers even after his retirement; an interesting feature of this research being that it was carried out by means of a kitchen tap and sieve. He also maintained a close interest in nuclear physics and it was in his laboratory that the Szilard–Chalmers process and the production of neutrons by x-rays in the photonuclear effect were discovered in 1934. But most of his interest lay in the fields of medical physics and radiology, in which he carried out a great deal of pioneering work. He played an important part in setting up the Strangeways Laboratory in Cambridge and in the installation of the million-volt x-ray machine at Barts, the first to be used for super-voltage therapy. In recent years he devoted a great deal of energy to the installation at Barts of the 15 Mev linear accelerator.

Apart from the direct contributions he has made to radiotherapy, he has been very active in the organization and coordination of cancer research. He was a founder-member of the British Empire Cancer Campaign, and until his death he was a member of its Grand Council and its Honorary Secretary. He was a past President and Silvanus Thompson Medallist of the British Institute of Radiology, Honorary Member of the Faculty of Radiologists, and for many years Secretary to the British Committee for Radiological Units.

All these outside activities did not detract from his devotion to Barts, which he served with the utmost loyalty and where he was loved and admired by students and colleagues. During the Second World War, when the Pre-Clinical School was evacuated to Queens' College, Cambridge, Hopwood was appointed Vice-Dean, a post he held until his retirement. He was mainly responsible for the excellent relationships that existed between the two Colleges during the seven years of evacuation, and in recognition of this the University of Cambridge conferred on him the Honorary Degree of M.A., and Queens' College elected him an Honorary Member.

A tribute to the memory of Professor Hopwood would not be complete without a reference to his remarkable personality. He was a big man both

physically and mentally. He had a great sense of humour and an inexhaustible store of anecdotes; he was a charming companion, tolerant, tactful and generous. He was always available to give advice, help and comfort, and there were many who leaned on his wisdom, experience and sympathy. He had a gift of infecting people with his own enthusiasm and vitality. Many of the leading authorities in medical physics are either his former pupils or people whom he introduced and inspired to enter into this new field. He managed to stimulate an interest in physics, and in science in general, in many thousands of medical students who passed through his hands. He will be greatly missed by his former students, his colleagues and his friends.

J. ROTBLAT.

JOHN EDWARD LENNARD-JONES

In a brief obituary notice to a specialized journal it is impossible to do justice to all the aspects of the varied life and work of John Edward Lennard-Jones, mathematician, physicist, theoretical chemist, civil servant and administrator. Nor indeed can space be found there for a full record of his achievements and of the distinctions he received in recognition of his services. A fuller biography will doubtless appear elsewhere.

Faced with the task of selection, I inevitably recall the years 1925 to 1932 when we were close and intimate colleagues, a period also when in the prime of early manhood, Lennard-Jones was maturing from a state of marked promise as a researcher and teacher to that of high distinction and reputation.

Coming to the University of Bristol from Manchester and Cambridge, where he had already attracted attention by his estimation of molecular fields from gaseous viscosity, from the equation of state and from crystal measurements, he here developed the pattern of so much of his subsequent work and career. The investigations he then carried out on the lattice properties of crystals and the processes of adsorption and diffusion on surfaces, in terms of interatomic forces, will find a permanent place in the literature. Moreover in this period he started to develop the molecular orbital technique for the study of diatomic molecules, which in its extension to more complex molecules has proved to be of so particular an interest to chemists. Indeed it is only three months ago that the eighteenth paper under the heading ' Molecular Orbital Theory of Chemical Valency ' was published from the Cambridge Laboratory where he held the Plummer Chair of Chemistry from 1932 until a year ago.

In the light of its importance to-day, it is interesting to recall the status of the subject of theoretical physics in 1925 when Bristol initiated a readership in it and invited Lennard-Jones to fill the appointment. Departments of Applied Mathematics were, of course, well known: but I am not aware of any permanent senior post at that time designed for a theorist working within the framework of a Physics Laboratory in a University of this country. Indeed, the very term ' theoretical physicist ' was not then in use in Britain.

Two years after his arrival, the Henry Herbert Wills Physical Laboratory was completed, and a chair, the first in this subject in the country, was created for him. By then he had already begun to throw himself with all his drive and enthusiasm into the project of creating in a provincial centre a new school of theorists and experimenters, trained together, and working in daily contact

and close co-operation. When he left for Cambridge five years later, having helped in the meantime to secure substantial endowments for the Wills Laboratory from the Rockefeller Foundation and from Mr. Melville Wills, who endowed his chair, his pioneering work in its formative years had already borne fruit. Moreover, by his example, the conception behind the scheme was beginning to catch on in other centres. Indeed, when he himself was invited in 1932 to Cambridge as the first theorist to fill the Plummer Chair, he was again chosen to break fresh ground, though this time in a department of Chemistry.

At Bristol also, I remember his early interest in administration, as a member of Senate and for two years as Dean of the Faculty of Science. He seemed even then to enjoy with others the task of framing policy ; and the work also gave opportunity for the exercise of his orderly mind in matters of detail and procedure.

With the advent of war it was therefore not surprising to find him by 1942 in high office as Chief Superintendent of the Armament Research Establishment, and later as Chief Scientific Officer and Director General of Scientific Defence Research in the Ministry of Supply.

But the war over and in Cambridge again, re-established in an atmosphere of active research and academic influence, he might have been expected to have remained content in such pleasant and friendly surroundings. It therefore required courage and initiative on his part to be prepared to take up the challenge of a new task by accepting an invitation to become Principal of the University College of Staffordshire at this early stage in its life. How sad it is to record his death at the early age of 60, only a year after he and his wife had taken up their residence there.

One final word : readers of the literature, seeking his papers prior to 1925, will find them under the heading J. E. Jones. On his marriage to Kathleen Mary Lennard, daughter of the late Alderman Lennard of Leicester, he changed his name by deed poll. Thus arose the familiar abbreviation L-J, by which he was so often addressed or spoken of, to the end of his life, by an ever increasing circle of friends and fellow workers.

A. M. TYNDALL.

ROBERT STEWART WHIPPLE

Robert Stewart Whipple, born on 1st August, 1871 was the son of G. M. Whipple, the Superintendent of Kew Observatory, and so was brought up in an environment in which scientific instruments played an important part. After leaving King's College School in 1888, he served as an Assistant in Kew Observatory under Dr. Charles Chree for eight years.

For a short time he was Assistant Manager to Mr. L. P. Casella, and two years later joined Sir Horace Darwin as a personal assistant. Mr. Horace Darwin, as he was then, had founded the Cambridge Scientific Instrument Company eighteen years earlier, and it was in connection with this Company that R. S. Whipple was to spend the remainder of his life. He was made Joint Managing Director in 1909 and served in this capacity up to his retirement in 1935. From then until 1949 he was the Chairman of the Board of Directors.

While his specialized early work was on the measurement of temperature, he was at all times interested in all types of scientific instruments. In a paper

on " Thermometers and Pyrometers " (1904) he says " one cannot help being struck with the interdependence of all branches of scientific work ". This consciousness was to remain with him all his life. The interest in thermometry was coupled with an early realization of the importance of temperature measurement in industry, and later his interest turned to the control of temperature in industry. Radiation pyrometry was little used in this country until Whipple introduced it, and his zeal caused him to press for its adoption in industry up and down the country. He secured the rights of manufacture in this country of the Féry pyrometer for the Cambridge Instrument Company. He invented the Whipple Temperature Indicator for use with electrical resistance thermometers, and it is still in considerable use.

While he was a first-class physicist, much of his success was due to his practical outlook. For example, in a paper on " Pyrometry as applied to the making of pottery " (1913), in assessing the requirements for a successful pyrometer installation for pottery making, he put them in the order (i) simplicity, (ii) durability, and (iii) accuracy. In the same paper he said " . . . it is essential to assure the fireman that the purpose of the installation is to help him in his work ".

This brings to mind his outstanding personal quality—a kindliness and gentleness of manner resulting from a sympathetic character. There are many young scientists who have been encouraged by a kind word from R. S. Whipple.

He served on many committees. He was President of the Optical Society, and was Treasurer of this Society for ten years, and a Vice-President for three years. He served for twenty-one years on the Board of the Institute of Physics, of which he was a Founder Member, and the Institute owes much to his generosity. He was one of the Founder Members of the Physical Society Club. He served the Royal Institution, as Visitor and Manager, and he was a Member of the Institution of Electrical Engineers, and served on its Council. He delivered the Faraday Lecture of that body in 1936. He was twice elected President of the Scientific Instrument Manufacturers' Association. He presided over Section A of the British Association in 1939.

His interest in all types of scientific instruments caused him to make a collection of historical scientific instruments and books; he gave this valuable collection to the University of Cambridge, together with a large sum of money for further purchases of books and apparatus. This formed the nucleus of a Museum of the History of Science, opened in 1951 as the " Whipple Museum ". It is in large measure due to him that the History and Philosophy of Science is now a subject for the Natural Sciences Tripos.

He served his locality also, taking much interest in the Highgate Literary and Scientific Institution of which he was President for 16 years, and one of his last acts just before his death was to found a trust for the advancement of the sciences and arts especially in North London. Here again his breadth of outlook was shown by the latitude he left to the trustees.

He was, as a foreigner once described him, " un homme plus que gentil ". To meet him was to respect him, to know him was to love him.

A. C. MENZIES.

MAURICE EDMOND JOSEPH GHEURY DE BRAY

We record with regret the death of M. E. J. Gheury de Bray who had been a fellow of the Physical Society from 1933 until he resigned through ill health in 1950. During the late 1940's Gheury de Bray compiled for his own use a comprehensive subject index of volumes 1–10 of the Physical Society's *Reports on Progress in Physics*. He presented the Physical Society with a copy of the Index, and this was published by the Society in 1953.

FRED HARRISON

We record with regret the death on 1st May of Fred Harrison, who had been a member of the Physical Society since 1893, and was thus one of the Society's oldest members.

CONTENTS OF SECTION A

PAGE

Mr. V. C. Officer and Mr. P. J. Eccles. The Time Coherence of Associated Cosmic Ray Particles 1037

Mr. H. R. Lemmer and Dr. M. A. Grace. The Angular Correlation of Successive γ-Rays in ^{60}Ni at Low Temperatures 1051

Mr. J. Hubbard. Plasma Oscillations in a Periodic Potential: The One-Zone Theory 1058

Dr. B. L. Moiseiwitsch and Dr. A. L. Stewart. Inelastic Collisions between Heavy Particles: III—Excitation of Helium Atoms in Fast Encounters with Hydrogen Atoms, Protons and Positive Helium Ions 1069

Dr. B. H. Bransden, Dr. A. Dalgarno and Mr. N. M. King. Electron Capture: IV—Capture from Helium Atoms by Fast Protons 1075

Dr. P. Swan. The Elastic Scattering of Electrons by the Excited 2s and 2p States of Atomic Hydrogen 1086

Dr. P. J. Campion. A Cloud Chamber Study of some Aspects of the Geiger Discharge 1095

Research Notes:

Dr. R. Huby. Phase of Matrix Elements in Nuclear Reactions and Radioactive Decay 1103

Mr. D. St. P. Bunbury. An Investigation of the (γ, n) Reaction in Cu, Zn and Ag 1106

Dr. G. W. Greenlees. Levels of ^{24}Mg from the ^{27}Al(p, α)^{24}Mg Reaction . 1107

Dr. E. Breitenberger. On Source Scattering in Angular Correlation Experiments with Soft Electrons 1108

Letters to the Editor:

Prof. M. H. L. Pryce. Nuclear Shell Structure 1111

Dr. K. M. Gatha, Mr. N. J. Patel and Mr. P. F. Patel. The Elastic Scattering of 125 mev Electrons by Beryllium 1111

Mr. P. Marin, Mr. G. R. Bishop and Dr. H. Halban. The Absolute Standardization of the 2·615 mev γ-rays of ThC″ and the Cross Section for the Photo Disintegration of the Deutron at this Energy 1113

Reviews of Books 1114

Obituary Notices:

Herbert Stanley Allen 1117
Alfred Blackie 1118
Albert Campbell 1119
Francis Cecil Chalklin 1122
Reginald Stanley Clay 1124
Frank Kenneth Goward 1126
Frank Lloyd Hopwood 1127
John Edward Lennard-Jones 1128
Robert Stewart Whipple 1129
Maurice Edmond Joseph Gheury de Bray 1131
Fred Harrison 1131

Contents of Section B 1132

Subject Index, Section A, Vol. 67 1133

Index of Authors (with Titles), Section A, Vol. 67 1141

Index to Reviews of Books, Section A, Vol. 67 1151

PROCEEDINGS OF THE PHYSICAL SOCIETY
SECTION B, 1954—VOL. 67

SUBJECT INDEX

PAGE

Aberrations, field, in wide aperture optical systems 801
Aberrations, primary, of anamorphotic lens systems 523
Aberrations, third-order, of doubly symmetric systems 523
Abrasion of rubber 883
Absorption of sound, *see under* Sound.
Accelerating fields, oxide cathodes in (R) 439
Adiabatic theory of electron gun for crossed field devices 449
Air stream, naturally convected, temperature decay law 42
Alcomax III, anhysteretic magnetization (R) 584
Alloy, Cu–Pd, variation of degree of order with temperature, experimental study . 297
Alloy, precipitation hardened, mechanisms of creep 832
Alloys, Pd–Ag, elastic constants and density (L) 266
Alloys, Pt–Co, intrinsic magnetization 38
Alloys, tin-rich, binary, lattice spacings 737
Aluminium, behaviour when deformed under alternating stresses (R) . . . 260
Anisotropic media, uniformly moving dislocations in 615
Anisotropic strata, transient heat flow in 731
Astro-optics, testing, application of shearing interferometer with fixed shear . . 232
Atmosphere, non-uniformly recombinant, monochromatically ionized layer in . . 717

Beta-ray spectrometer, *see* Spectrometer, β-ray.
Bismuth, infra-red properties (L) 269
Breakdown, form of, in germanium diodes 328

Calcium, vapour pressure : I 783
Calcium, vapour pressure : II 787
Capillary flow of liquids, meniscus resistance in 395
Carbon dioxide in narrow tubes, absorption of sound in 65
Cathode sputtering in glow discharges 546
Cathodes, oxide-coated, *see* Oxide-coated cathodes.
Cathodo-luminescence of MgO 811
Chapman's grazing incidence integral Ch(x, χ), tables 304
Circumscribed liquid, surface vibrations of (R) 899
Cobalt, crystal structure, effect of grain size 456
Cobaltite, properties of point contacts on (L) 589
Coercivity of iron oxide powders, effect of particle shape variation 875
Colour centres in quartz, formation (R) 262
Colour equations, additivity : II 513
Colour vision, normal, and practical colour matching 477
Computing machine, electronic, Manchester University, ray tracing on . . . 569
Conductivity, electrical and thermal, of semiconductors (R) . . . 360, corr. 599
Constrictions and jogs in extended dislocations 427
Contact area, true, between solids 309
Contact, injection, ratio and minority carrier lifetime of transistor materials . . 9
Contacts, metallic, effect of compressive and shearing forces on surface films present in 238
Contacts, metal–semiconductor, theory of rectification and injection at . . . 575
Contacts, point, on cobaltite, properties (L) 589
Contacts, point, on Ge, effect of temperature gradients (L) 651
Counters, Geiger–Muller, self-quenching, velocity of discharge propagation in (R) . 161
Counter, x-ray diffraction, gas-flow (R) 357
Cracks, ring, *see* Ring cracks.

	PAGE
Creep, influence on measured hardness of soft metals	563
Creep, mechanisms, in precipitation hardened alloy	832
Crystals, organic, fluorescence spectra	657
Crystals, scintillating, luminescence characteristics	817
Crystals, *see also* Germanium.	
Crystal structure of cobalt, effect of grain size	456
Current gain of germanium filamentary transistors	625
Current multiplication processes in n-type germanium point-contact transistors	636
Damping measurement, determination of relaxation spectra from	338
Deformation of aluminium under alternating stresses (R)	260
Diamond, formation of trigons on	538
Diamond, oriented ring cracks on	473
Dielectric constant of isotropic powders, method of measuring (L)	267
Dielectric relaxation of polar molecules in solution	149
Diffraction patterns, our-of-focus, for microscope objectives	825
Diffraction, *see also* X-ray, Electron.	
Diffusion and mass transport in tubes (38th Guthrie Lecture)	857
Diffusion, thermal, gaseous, approach to steady state, and application to determination of dependence of gas diffusivity on concentration ratio	753
Diodes, germanium, *see* Germanium, diodes.	
Discharge propagation velocity in self-quenching Geiger–Müller counters (R)	161
Discharge, *see also* Glow discharges.	
Dislocations, extended, constrictions and jogs in	427
Dislocations, uniformly moving, in anisotropic media	615
Dissociative recombination at surfaces (L)	495
Drift mobility of injected holes in germanium, temperature dependence of	18
Dynamic mechanical properties of organic glasses	120
Editorial	1
Elastic and piezoelectric coefficients of dipotassium tartrate (DKT)	422
Elastic solid, friction of	89
Electrical conductivity and other properties of coating in oxide-coated cathodes, measurements	387
Electrical conductivity of some condensed aromatic hydrocarbons (R)	892
Electrical conductivity, *see also under* Conductivity.	
Electron beam loading in linear accelerators, theory	705
Electronegativities, scale, and semiconducting compounds (R)	258
Electron gun for crossed field devices, adiabatic theory	449
Electrons, mobility, mean free path varying with velocity (R)	159, corr. 279
Electrostatic forces on immersed bodies, theory	689
Emission, secondary, *see* Secondary emission.	
Emulsions, photographic, *see* Photographic emulsions.	
Excitation, cascade, of phosphors, by ultra-violet radiation (L)	496
Fading, short period, in ionospheric turbulence, mechanism and distribution	211
Fields, electric and magnetic, crossed, electron gun for, adiabatic theory	449
Filaments, germanium, noise at very low frequencies (L)	271
Films, metal evaporated, cause of stress in (L)	497
Films, surface, *see* Surface films.	
Fluids, superposed, Taylor instability of interface, solution by successive approximations	748
Fluorescence spectra of organic crystals	657
Forces, compressive and shearing, effect on surface films present in metallic contacts	238
Formation current in electrolytic oxidation of zirconium	201
Friction of an elastic solid	89

PAGE

Gas diffusivity, dependence on concentration ratio, and approach to steady state in gaseous thermal diffusion 753
Germanium diodes, breakdown in, observation of 328
Germanium, effect of temperature gradients at point contacts on (L) 651
Germanium filamentary transistors, current gain of 625
Germanium filaments, noise at very low frequencies (L) 271
Germanium, n-type, point-contact transistors, current multiplication processes in . 636
Germanium, single crystal, interferometric observation of mosaic structure on (111) face (R) 359
Germanium, single crystal, relation between sum of donor and acceptor concentration and lifetime (R) 646
Germanium, surface properties, measurement 409
Germanium, temperature dependence of drift mobility of injected holes in . . 18
Glass, oriented ring cracks on 467
Glasses, organic, dynamic mechanical properties 120
Glow discharges, cathode sputtering in 546
Grain growth in metals, mechanism, with special reference to steel 131
Grain size, effect on crystal structure of cobalt 456
Grazing incidence integral $Ch(x, \chi)$, tables 304
Guthrie lecture, 38th—Diffusion and mass transport in tubes 857

Hardness, measured, of soft metals, influence of creep 563
Hardness scale, Mohs's, physical interpretation 249
Heat flow, effects on thermoelectric power in semiconductors (L) 587
Heat flow, transient, in anisotropic strata 731
Heat flow in Transvaal and Orange Free State 664
Heat-treatment and quenching from high temperatures of small specimens in vacuum, apparatus for (R) 845
Helium temperatures, resistance–temperature characteristics of leaded brass wires (R) 644
Hydrocarbons, aromatic, condensed, electrical conductivity (R) 892

Imaging systems, doubly symmetric, third-order aberrations 523
Immersed bodies, theory of electrostatic forces on 689
Impact of spheres on plates . . . . . 677, corr. 794
Indium antimonide, interpretation of properties 775
Indium antimonide, optical and photoelectric properties 761
Infra-red properties of bismuth (L) 269
Infra-red radiation, phosphors emitting 442
Injected holes in germanium, temperature dependence of drift mobility . . 18
Injection and rectification at metal–semiconductor contact, theory . . . 575
Injection, contact, *see* Contact injection.
Interference microscopy, *see* Microscopy, interference.
Interferometer, rotating, for the measurement of direction of arrival of short radio waves 553
Interferometer, shearing, with fixed shear, and application to testing of astro-optics . 232
Interferometric observation of mosaic structure on (111) face of single crystal of germanium (R) 359
Ionized layer, monochromatic, in non-uniformly recombinant atmosphere . . 717
Ionosphere, D and E regions, application of characteristics of monochromatically ionized layer in non-uniformly recombinant atmosphere 717
Ionospheric turbulence, mechanism and distribution of short period fading in . . 211
Iron oxide powders, effect of particle shape variation on coercivity 875

Jet noise, choked, reduction of 313
Jogs and constrictions in extended dislocations 427

PAGE
Lattice spacings of binary tin-rich alloys 737
Lead sulphide, selenide and telluride, measurements on optical properties . . 2
Lens systems, anamorphotic, primary aberrations 529
Light, scattering, in photographic emulsion layers (R) 164
Light, totally reflected, polarization 608
Lightning flashes, comparison of rates of change of current in step and return processes (R) 582
Linear accelerators, electron beam loading, theory 705
Low temperatures, *see also* Helium temperatures.
Luminescence characteristics of some scintillating crystals 817
Luminescence, *see also* Cathodo-luminescence, Phosphors.

Magnetic fields, intense, effect on electroluminescent powder phosphors. . . 870
Magnetic lens spectrometer, *see under* Spectrometer.
Magnetic susceptibility of metallic uranium 28
Magnetic transitions and structural changes in hexagonal manganese compounds . 290
Magnetization, anhysteretic, of Alcomax III (R) 584
Magnetization, intrinsic, in Pt–Co alloys 38
Manganese compounds, hexagonal, magnetic transitions and structural changes . 290
Mass transport and diffusion in tubes (38th Guthrie lecture). 857
Membranes, square, vibrating, symmetry of (R). 895
Mercury arc, low-pressure, current limitation in. 768
Metal films, evaporated, cause of stress in (L) 497
Metals, mechanism of grain growth in, with special reference to steel . . . 131
Metals, soft, influence of creep on measured hardness 563
Metal surfaces, irregular, work function of. 187
Micro-method for measurement of self-diffusion coefficients of solutions. . . 401
Microscope objectives, out-of-focus diffraction patterns for 825
Microscopy, interference, for determination of refractive indices of inhomogeneous solids 138
Microscopy, interference, at high wedge angles 281
Minority carrier lifetime and contact injection ratio of transistor materials . . 9
Mobility of electrons with mean free path varying with velocity, formula (R) 159, corr. 279
Mohs's hardness scale—a physical interpretation. 249

Negative ion emission from oxide-coated cathodes : I 169
Negative ion emission from oxide-coated cathodes : II. 177
Negative ion formation in CCl_4 and $TiCl_4$ (R) 439
Newton's rings of high visibility (R) 492
Noise, choked jet, reduction of 313
Noise in germanium filaments at very low frequencies (L) 271

Obituaries 905
Optical and photoelectric properties of indium antimonide 761
Optical properties of lead sulphide, selenide and telluride, further measurements . 2
Order, degree of, variation with temperature in Cu–Pd alloy, experimental study . 297
Oxidation, electrolytic, of zirconium, analysis of formation current in . . . 201
Oxide cathodes in accelerating fields (R) 439
Oxide cathodes, reactions with gases at very low pressures 369
Oxide-coated cathodes, coating, properties including electrical conductivity and thermoelectric power 387
Oxide-coated cathodes, negative ion emission from : I 169
Oxide-coated cathodes, negative ion emission from : II 177

Phosphors, cascade excitation by ultra-violet radiation (L) 496
Phosphors emitting infra-red radiation (L) 442
Phosphors, powder, electroluminescent, effect of intense magnetic fields. . . 870

PAGE
Photoelectric and optical properties of indium antimonide 761
Photographic emulsion layers, scattering of light in (R) 164
Photographic emulsions, unexposed, scattering effects in, application of radiative transfer theory to 105
Physics in the university and the nation (Presidential Address) 601
Piezoelectric and elastic coefficients of dipotassium tartrate (DKT) 422
Point contacts, *see* Contacts, point.
Polarization of totally reflected light 608
Polar molecules in solution, dielectric relaxation 149
Positive column, theory 485
Powders, isotropic, dielectric constant, method of measuring (L) 267
Presidential Address—Physics in the university and the nation 601
Psychrometry, thermocouple, error and accuracy in 217

Quartz, formation of colour centres in (R) 262
Quenching from high temperatures, and heat-treatment of small specimens in vacuum, apparatus for (R) 845

Radiative, infra-red, *see* Infra-red radiation.
Radiation, ultra-violet, *see* Ultra-violet radiation.
Radio waves, short, measurement of direction of arrival by rotating interferometer 553
Ray tracing on Manchester University electronic computing machine . . . 569
Recombination, dissociative, at surfaces (L) 495
Recombination, statistics of populations with 227
Rectification and injection at metal–semiconductor contact, theory 575
Refractive indices of inhomogeneous solids, determination, by interference microscopy 138
Relaxation spectra, determination from damping measurements 338
Resistance–temperature characteristics of leaded brass wires at helium temperatures (R) 644
Ring cracks on diamond, oriented 473
Ring cracks on glass, optical studies 467
Rubber, abrasion 883

Scattering effects in unexposed photographic emulsions, application of radiative transfer theory 105
Scattering of light in photographic emulsion layers (R) 164
Scintillating crystals, luminescence characteristics 817
Secondary emission, electron, of sodium and zinc (R) 843
Self-diffusion coefficients of solutions, micro-method for measurement . . . 401
Semiconducting compounds and scale of electronegativities (R) 258
Semiconducting compounds with zinc-blende structure, measurements on (L) . . 848
Semiconductors, effects of heat flow on thermoelectric power in (L) 587
Semiconductors, thermal and electrical conductivity (R) . . . 360, corr. 599
Slip, effect on sound propagation (L) 265
Sodium, secondary electron emission (R) 843
Solids, inhomogeneous, refractive index determination by interference microscopy . 138
Solids, true contact area between 309
Solutions, self-diffusion coefficients, micro-method for measurement . . . 401
Sound, absorption in carbon dioxide contained in narrow tubes 65
Sound, absorption, and velocity, in aqueous sugar solutions 70
Sound propagation, effect of slip on (L) 265
South Africa, heat flow in Transvaal and Orange Free State 664
Spectra, fluorescence, *see* Fluorescence.
Spectra, relaxation, *see* Relaxation spectra.
Spectrometer, β-ray, double magnetic lens, design and operation 52
Spectrometer, vacuum, Geiger counter, for study of soft x-ray lines 348

PAGE

Statistics of a population with creation and recombination dependent on existing numbers 227
Steel, grain growth in, mechanism 131
Strain ageing and discontinuous flow in zinc crystals 98
Stress in evaporated metal films, cause (L) 497
Structural changes and magnetic transitions in hexagonal manganese compounds . 290
Sugar solutions, aqueous, absorption and velocity of sound in 70
Surface films present in metallic contacts, effect of compressive and shearing forces . 238
Surface properties of germanium, measurement. 409
Surfaces, dissociative recombination at (L) 495
Susceptibility of tantalum, temperature variation 728

Tantalum, susceptibility, temperature variation 728
Taylor instability of interface between superposed fluids, solution by successive approximations 748
Teaching of physics, and the nation (Presidental Address) 601
Temperature decay law of naturally convected air stream 42
Temperature gradients at point contacts on germanium, effect (L) 651
Temperature–resistance characteristics of leaded brass wires at helium temperatures (R) 644
Temperature variation of susceptibility of tantalum 728
Thermal conductivity, *see also under* Conductivity.
Thermal diffusion, *see* Diffusion.
Thermoelectric power and other properties of coating in oxide-coated cathodes, measurements 387
Transfer theory, radiative, application to scattering effects in unexposed photographic emulsions 105
Transistors, germanium, filamentary, current gain of 625
Transistors, germanium, n-type, point-contact, current multiplication processes in . 636
Transistor materials, minority carrier lifetime and contact injection radio. . . . 9
Trigons, formation on diamonds 538

Ultra-violet radiation, cascade excitation of phosphors (L) 496
Uranium, metallic, magnetic susceptibility. 28

Vapour pressure of calcium : I 783
Vapour pressure of calcium : II 787
Vibrating square membrane, symmetry of (R) 895
Vibrations, surface, of a circumscribed liquid (R) 899
Viscometry—meniscus resistance in capillary flow of liquids 395
Vision, colour, *see* Colour vision.
Visual threshold, order of coincidence required 673

Wires, leaded brass, resistance–temperature characteristics at helium temperatures (R) 644
Work function of irregular metal surfaces 187

X-ray diffraction counter, gas-flow (R) 357
X-rays, soft, study of lines by Geiger counter vacuum spectrometer 348

Zinc-blende structure semiconducting compounds, measurements on (L) . . . 848
Zinc crystals, discontinuous flow, and strain ageing 98
Zinc, secondary electron emission (R) 843
Zirconium, analysis of formation current in electrolytic oxidation 201

INDEX OF AUTHORS (WITH TITLES)

PAGE

Allen, J. E., and Thonemann, P. C. : Current limitation in the low-pressure mercury arc 768
Arndt, U. W., Coates, W. A., and Crathorn, A. R. : A gas-flow x-ray diffraction counter (R) 357
Avery, D. G. : Further measurements on the optical properties of lead sulphide, selenide and telluride 2
Avery, D. G., Goodwin, D. W., Lawson, W. D., and Moss, T. S. : Optical and photo-electrical properties of indium antimonide 761
Ayers, S., with Bechmann, R. : Elastic and piezoelectric coefficients of dipotassium tartrate (DKT) 422

Bates, L. F., and Hughes, D. : The magnetic susceptibility of metallic uranium . 28
Bechmann, R., and Ayers, S. : Elastic and Piezoelectric coefficients of dipotassium tartrate (DKT) 422
Bell, D. A. : Statistics of a population with creation and recombination dependent on existing numbers 227
Benbow, J. J. : The dynamic mechanical properties of some organic glasses . . 120
Bilby, B. A., with Bullogh, R. : Uniformly moving dislocations in anisotropic media . 615
Birks, J. B., and Wright, G. T. : Fluorescence spectra of organic crystals . . 657
Black, G. : Ray tracing on the Manchester University electronic computing machine 569
Blaschke, W. S. S. : Field aberrations in wide aperture optical systems . . . 801
Brindley, G. S. : The order of coincidence required for visual threshold . . . 673
Brown, D. : A shearing interferometer with fixed shear and its application to some problems in the testing of astro-optics 232
Bullough, R., and Bilby, B. A. : Uniformly moving dislocations in anisotropic media . 615
Burfoot, J. C. : Third-order aberrations of ' doubly symmetric ' systems . . 523

Cade, R. : Contributions to the theory of electrostatic forces on immersed bodies . 689
Calderwood, G. F. N., and Mardles, E. W. J. : Viscometry—the meniscus resistance in capillary flow of liquids 395
Carte, A. E. : Heat flow in the Transvaal and Orange Free State 664
Chalklin, the late F. C., with Rogers, J. L. : A Geiger counter vacuum spectrometer and its use for the study of soft x-ray lines 348
Chapman, Sydney : A monochromatically ionized layer in a non-uniformly recombinant atmosphere ; with applications to the D and E ionospheric regions . 717
Charlesby, A., with Polling, J. J. : Analysis of the formation current in electrolytic oxidation of zirconium 201
Coates, W. A., with Arndt, U. W., and Crathorn, A. R. : A gas-flow x-ray diffraction counter (R) 357
Cocks, M. : The effect of compressive and shearing forces on the surface films present in metallic contacts 238
Cole, D. G., Feltham, P., and Gillam, E. : On the mechanism of grain growth in metals, with special reference to steel 131
Cook, J. R., and Mahmoud, K. A. : Luminescence characteristics of some scintillating crystals 817
Craggs, J. D., with Marriott, J., and Thorburn, R. : Negative ion formation in CCl_4 and $TiCl_4$ 437
Crathorn, A. R., with Arndt, U. W., and Coates, W. A. : A gas-flow x-ray diffraction counter (R) 357
Crittenden, E. C., Jr., with Hoffman, R. W., and Daniels, R. D. : The cause of stress in evaporated metal films (L) 497
Cunnell, F. A., Edmond, J. T., and Richards, J. L. : Measurements on some semiconducting compounds with the zinc-blende structure (L) 848

PAGE

Daniels, R. D., with Hoffman, R. W., and Crittenden, E. C., Jr. : The cause of stress in evaporated metal films (L) 497
Dain, J., and Lewis, I. A. D. : Adiabatic theory of an electron gun for crossed field devices 449
Davidson, P. M. : A note on the formula for the mobility of electrons with mean free path varying with velocity (R) . . . . . 159, corr. 279
Daykin, P. N., and Wright, D. A. : Oxide cathodes in accelerating fields (R) . . 439
Douglas, P. E. : The vapour pressure of calcium : I 783
Dumbleton, M. J. : Discontinuous flow in zinc crystals and its relationship to strain ageing 98
Dumbleton, M. J., with Garlick, G. F. J. : Phosphors emitting infra-red radiation (L) 442
Dyson, J., and Hirst, W. : The true contact area between solids 309

Ecker, G. : Theory of the positive column. 485
Ecker, G., and Emeléus, K. G. : Cathode sputtering in glow discharges 546
Edmond, J. T., with Cunnell, F. A., and Richards, J. L. : Measurements on some semiconducting compounds with the zinc-blende structure (L) 848
Ellis, J., with Rundle, G. P., Griffith, T. C., and Tomlinson, H. S. : The design and operation of a double magnetic lens beta-ray spectrometer 52
Emeléus, K. G. : Dissociative recombination at surfaces (L). 495
Emeléus, K. G., with Ecker, G. : Cathode sputtering in glow discharges. . . . 546
Enomoto, H., with Minozuma, F. : The mechanism and distribution of short period fading under conditions of ionospheric turbulence 211

Farnell, G. C. : Scattering of light in photographic emulsion layers (R) 164
Faust, R. C. : The determination of the refractive indices of inhomogeneous solids by interference microscopy 138
Feltham, P., with Cole, D. G., and Gillam E. : On the mechanisms of grain growth in metals, with special reference to steel 131

Gandy, R. O. : Out-of-focus diffraction patterns for microscope objectives . . . 825
Garlick, G. F. J., and Dumbleton, M. J. : Phosphors emitting infra-red radiation (L) 442
Gibson, A. F., with Lawrance, R., and Granville, J. W. : On the current gain of germanium filamentary transistors 625
Gillam, E., with Cole, D. G., and Feltham, P. : On the mechanism of grain growth in metals, with special reference to steel 131
Goldsmid, H. J. : On the thermal and electrical conductivity of semiconductors (R) . 360, corr. 599
Goodman, C. H. L. : Semiconducting compounds and the scale of electronegativities (R) 258
Goodwin, D. W., with Avery, D. G., Lawson, W. D., and Moss, T. S. : Optical and photoelectrical properties of indium antimonide 761
Gould, J. E., and McCaig, M. : Anhysteretic magnetization of alcomax III (R) . 584
Granville, J. W., with Lawrance, R., and Gibson, A. F. : On the current gain of germanium filamentary transistors 625
Grattidge, W., and Shepherd, A. A. : Negative ion emission from oxide-coated cathodes : II 177
Griffith, T. C., with Rundle, G. P., Ellis, J., and Tomlinson, H. S. : The design and operation of a double magnetic lens beta-ray spectrometer 52
Gunn, J. B. : Measurement of the surface properties of germanium 409
Gunn, J. B. : The theory of rectification and injection at a metal–semiconductor contact 575

Halperin, A. : The formation of trigons on diamonds 538
Halperin, A., with Tolansky, S. : Oriented ring cracks on diamond 473
Hatherley, M., with Louat, N. : The behaviour of aluminium deformed under alternating stresses 260
Henderson, S. T., with Ranby, P. W. : Cascade excitation of phosphors by ultra-violet radiation (L) 496

PAGE

Henisch, H. K., and Tipple, P. M. : A note on the effect of temperature gradients at point contacts on germanium (L) 651
Hill, Nora E. : The dielectric relaxation of polar molecules in solution 149
Hirst, W., with Dyson, J. : The true contact area between solids 309
Hoare, F. E., Kouvelites, J. S., Matthews, J. C., and Preston, J. : The temperature variation of susceptibility of tantalum (R) 728
Hoare, F. E., and Yates, B. : The elastic constants and density of palladium silver alloys (L) 266
Hodges, D. B. : A comparison of the rates of change of current in the step and return processes of lightning flashes (R) 582
Hodgson, J. N. : The infra-red properties of bismuth 269
Hoffman, R. W., Daniels, R. D., and Crittenden, E. C., Jr. : The cause of stress in evaporated metal films (L) 497
Hogarth, C. A. : Current multiplication processes in n-type germanium point-contact transistors 636
Howell, H. G., with Lodge, A. S. : Friction of an elastic solid 89
Howes, V. R., with Tolansky, S. : Optical studies of ring cracks on glass. . . . 467
Hughes, D., with Bates, L. F. : The magnetic susceptibility of metallic uranium . 28

Ince, A. N. : The effect of intense magnetic fields on electroluminescent powder phosphors 870
Ingraham, R. L. : Taylor instability of the interface between superposed fluids—Solution by successive approximation. 748

Jones, D. Madoc, and Owen, E. A. : Experimental study of the variation of the degree of order with temperature in a copper–palladium alloy 297
Jones, D. Madoc, with Owen, E. A. : Effect of grain size on the crystal structure of cobalt 456

Kouvelites, J. S., with Hoare, F. E., Matthews, J. C., and Preston, J. : The temperature variation of susceptibility of tantalum 728

Lawley, L. E. : The absorption of sound in carbon dioxide contained in narrow tubes 65
Lawrance, R. : The temperature dependence of the drift mobility of injected holes in germanium 18
Lawrance, R., Gibson, A. F., and Granville, J. W. : On the current gain of germanium filamentary transistors 625
Lawson, W. D., with Avery, D. G., Goodwin, D. W., and Moss, T. S. : Optical photoelectrical properties of indium antimonide 761
Lee, J. A., and Raynor, G. V. : The lattice spacings of binary tin-rich alloys . . 737
Lempicki, A., and Wood, C. : Observations on a form of breakdown in germanium diodes 328
Lewis, I. A. D., with Dain, J. : Adiabatic theory of an electron gun for crossed field devices 449
Lewis, T. J. : The work function of irregular metal surfaces. 187
Lipson, H., and Cochran, W. : The determination of crystal structures 655
Lodge, A. S., and Howell, H. G. : Friction of an elastic solid 89
Louat, N., and Hatherley, M. : The behaviour of aluminium deformed under alternating stresses (R) 260
Lovett, C. M. : Some measurements of the electrical conductivity, thermoelectric power and other properties of the coating in the oxide-coated cathode . . 387
Low, G. G. E. : Properties of point contacts on cobaltite (L) 589

Mahmoud, K. A., with Cook, J. R. : Luminescence characteristics of scintillating crystals 817
Many, A. : Measurement of minority carrier lifetime and contact injection ratio on transistor materials 9

PAGE

Mardles, W. E. J., with Calderwood, G. F. N.: Viscometry—the meniscus resistance in capillary flow of liquids 395
Marriott, J., Thorburn, R., and Craggs, J. D.: Negative ion formation in CCl_4 and $TiCl_4$ (R) 437
Matthews, J. C., with Hoare, F. E., Kouvelites, J. S., and Preston, J.: The temperature variation of susceptibility of tantalum 728
Minozuma, F., and Enomoto, H.: The mechanism and distribution of short period fading under conditions of ionospheric turbulence 211
McCaig, M., with Gould, J. E.: Anhysteretic magnetization of alcomax III (R) . 584
Mitchell, E. W. J., and Paige, E. G. S.: On the formation of colour centres in quartz (R) 262
Monteith, J. L.: Error and accuracy in thermocouple psychrometry 217
Mortier, P. A. C., and Roose, J. F.: Velocity of discharge propagation in self-quenching Geiger–Müller counters (R) 161
Moss, T. S.: The interpretation of the properties of indium antimonide. . . . 775
Moss, T. S., with Avery, D. G., Goodwin, D. W., and Lawson, W. D.: Optical and photoelectrical properties of indium antimonide 761
Mykura, H.: Interference microscopy at high wedge angles 281

Nettley, P. T.: The approach to the steady state in gaseous thermal diffusion and its application to determining the dependence of gas diffusivity on the concentration ratio 753
Northrop, D. C., and Simpson, O.: Electrical conductivity of some condensed aromatic hydrocarbons (R) 892

Olds, G. C. E.: Mechanisms of creep in a precipitation hardened alloy 832
Orchard, G. A. J., with Ditchburn, R. W.: The polarization of totally reflected light. 608
Osmond, W. P.: The effect of particle shape variations on the coercivity of iron oxide powders 875
Owen, E. A., and Jones, D. Madoc: Effect of grain size on the crystal structure of cobalt 456
Owen, E. A., with Jones, D. Madoc: Experimental study of the variation of the degree of order with temperature in a copper–palladium alloy. 297

Paige, E. G. S., with Mitchell, E. W. J.: On the formation of colour centres in quartz (R) 262
Pant, L. M., with Saksena, B. D.: Cathodo-luminescence of MgO 811
Parkinson, D. H., and Quarrington, J. E.: The resistance–temperature characteristics of leaded brass wires at helium temperature (R) 644
Parrott, J. E.: The effects of heat flow on thermoelectric power in semiconductors (L) 587
Pearson, J. R. A., with Roesler, F. C.: Determination of relaxation spectra from damping measurements 338
Pitts, E.: The application of radiative transfer theory to scattering effects in unexposed photographic emulsions 105
Polling, J. J., and Charlesby A.: Analysis of the formation current in electrolytic oxidation of zirconium 201
Powell, A.: The reduction of choked jet noise 313
Preston, J., with Hoare, F. E., Kouvelites, J. S., and Matthews, J. C.: The temperature variation of susceptibility of tantalum 728
Pryor, A. W., and Roscoe, R.: The velocity and absorption of sound in aqueous sugar solutions 70

Quarrington, J. E., with Parkinson, D. H.: The resistance–temperature characteristics of leaded brass wires at helium temperatures 644

PAGE

Railston, W. : The temperature decay law of a naturally convected air stream . . 42
Ranby, P. W., and Henderson, S. T. : Cascade excitation of phosphors by ultra-violet radiation (L) 496
Ransom, P. and Rose, F. W. G. : On the relation between the sum of donor and acceptor concentration and lifetime in single crystal germanium (R). . . . 646
Raynor, G. V., with Lee, J. A. : The lattice spacings of binary tin-rich alloys . . 737
Reynolds, J. A. : A method of measuring the dielectric constant of isotropic powders (L) 267
Richards, D. A. : Newton's rings of high visibility (R) 492
Richards, J. L., with Cunnell, F. A., and Edmond, J. T. : Measurements on some semi-conducting compounds with the zinc-blende structure (L) 848
Roesler, F. C., and Pearson, J. R. A. : Determination of relaxation spectra from damping measurements 338
Rogers, J. L., and Chalklin, the late, F. C. : A Geiger counter vacuum spectrometer and its use for the study of soft x-ray lines 348
Rolfe, J. : A micro-method for the measurement of self-diffusion coefficients of solutions 401
Rollin, B. V., and Templeton, I. M. : Noise in germanium filaments at very low frequencies (L) 271
Rooksby, H. P., with Willis, B. T. M. : Magnetic transitions and structural changes in hexagonal manganese compounds 290
Roose, J. F., with Mortier, P. A. C. : Velocity of discharge propagation in self-quenching Geiger–Müller counters (R) 161
Roscoe, R., with Pryor, A. W. : The velocity and absorption of sound in aqueous sugar solutions 70
Rose, F. W. G., with Ransom, P. : On the relation between the sum of donor and acceptor concentration and lifetime in single crystal germanium 646
Rubenstein, C. : The influence of creep on the measured hardness of soft metals . 563
Rundle, G. P., Ellis, J., Griffith, T. C., and Tomlinson, H. S. : The design and operation of double magnetic lens beta-ray spectrometer 52

Saksena, B. D., and Pant, L. M. : Cathodo-luminescence of MgO 811
Saxon, G. : A theory of electron beam loading in linear accelerators . . . 705
Schallamach, A. : On the abrasion of rubber 883
Schofield, T. H. : An apparatus for heat-treatment and quenching small specimens in a vacuum from high temperatures (R) 845
Scott, D. R. : Transient heat flow in anisotropic strata (L) 731
Shepherd, A. A., with Grattidge, W. : Negative ion emission from oxide-coated cathodes : II 177
Simpson, A. W., and Tredgold, R. H. : Intrinsic magnetization in platinum cobalt alloys 38
Simpson, O., with Northrop, D. C. : Electrical conductivity of some condensed aromatic hydrocarbons 892
Stroh, A. N. : Constrictions and jogs in extended dislocations 427

Tabor, D. : Mohs's hardness scale—a physical interpretation 249
Taylor, G. I. : Diffusion and mass transport in tubes (38th Guthrie Lecture) . 857
Templeton, I. M., and Rollin, B. V. : Noise in germanium filaments at very low frequencies (L) 271
Thonemann, P. C., and Allen, J. E. : Current limitation in the low-pressure mercury arc 1
Thorburn, R., with Marriott, J., and Craggs, J. D. : Negative ion formation in CCl_4 and $TiCl_4$ (R) 437
Tillett, J. P. A. : A study of the impact of spheres on plates . . 677, corr. 794
Tolansky, S., and Halperin, A. : Oriented ring cracks on diamond 473
Tolansky, S., and Howes, V. R. : Optical studies of ring cracks on glass . . . 467
Tomlin, D. H. : The vapour pressure of calcium : II 787

PAGE

Tomlinson, H. S., with Rundle, G. P., Ellis, J., and Griffith, T. C.: The design and operation of a double magnetic lens beta-ray spectrometer 52
Tredgold, R. H., with Simpson, A. W.: Intrinsic magnetization in platinum–cobalt alloys 38
Trezona, P. W.: Additivity of colour equations: II 513

Verma, Ajit Ram: Interferometric observation of mosaic structure on the (111) face of a single crystal of germanium (R) 359
Vick, F. A., and Walley, C. A.: Negative ion emission from oxide-coated cathodes: I 169

Wagener, S.: Reactions between oxide cathodes and gases at very low pressures . 369
Waller, Mary D.: Note on surface vibrations of a circumscribed liquid (R) . . 899
Waller, Mary D.: Symmetry of vibrating square membrane (R) 895
Walley, C. A., with Vick, F. A.: Negative ion emission from oxide-coated cathodes: I 169
Warburton, F. L.: Variations in normal colour vision in relation to practical colour matching 477
Weston, D. E.: The effect of slip on sound propagation (L) 265
Whale, H. A.: A rotating interferometer for the measurement of the directions of arrival of short radio waves 553
Whiddington, R.: Physics in the university and the nation 601
Wilkes, M. V.: A table of Chapman's grazing incidence integral Ch(x, χ) . . 304
Willis, B. T. M., and Rooksby, H. P.: Magnetic transitions and structural changes in hexagonal manganese compounds 290
Wood, C., with Lempicki, A.: Observations on a form of breakdown in germanium diodes 328
Woods, J.: The secondary electron emission of sodium and zinc (R) 843
Wright, D. A., with Daykin, P. N.: Oxide cathodes in accelerating fields (R) . 439
Wright, G. T., with Birks, J.: Fluorescence spectra of organic crystals . . . 657
Wynne, C. G.: The primary aberrations of anamorphotic lens systems . . . 529

Yates, B., with Hoare, F. E.: The elastic constants and density of palladium–silver alloys (L) 266

INDEX TO REVIEWS OF BOOKS

PAGE

Alexander, P., and Hudson, R. F.: *Wool: its Chemistry and Physics* . . . 903
Allen, N. C. B., and Martin, L. H.: *Exercises in Experimental Physics* . . 82

Barkas, W. W., Hearmon, R. F. S., and Rance, H. F.: *Mechanical Properties of Wood and Paper* 444

De Barr, A. E.: *The Magnetic Circuit.* (Institute of Physics Monographs for Students) 274
De Barr, A. E.: *Soft Magnetic Materials Used in Industry.* (Institute of Physics Monographs for Students) 274

Beck, A. H. W.: *Thermionic Valves, their Theory and Design* 505
Bell, D. A.: *Statistical Methods in Electrical Engineering* 598
Bickley, W. G.: *Bessel Functions and Formulae* 593
Bowen, E. J., and Wokes, F.: *Fluorescence of Solutions* 366
Braddick, H. J. J.: *The Physics of Experimental Method* 853
Bronwell, A.: *Advanced Mathematics in Physics and Engineering* 84
Brown, J.: *Microwave Lenses* 502
Brown, R. C.: *A Textbook of Physics*, Vol. 3, *Sound* 900
Bruining, H.: *Physics and Applications of Secondary Electron Emission* . . . 853
Butler, J. A. V., and Randall, J. T. (Ed.): *Progress in Biophysics and Biophysical Chemistry*, Vol. III 277
Butschin, M. N.: *Reibungsuntersuchungen an geschmirten Oberflächen bei tiefen Temperaturen* 272
Byrd, P. F., and Friedman, M. D.: *Handbook of Elliptic Integrals for Engineers and Physicists* 850

Chalmers, B.: *Progress in Metal Physics*, Vol. 3 82
Chalmers, B., and King, R. (Ed.): *Progress in Metal Physics*, Vol. 5 . . . 901
Chaundy, T. W., Barrett, P. R., and Batey, C.: *The Printing of Mathematics* . 510
Copper Development Association: *Copper in Instrumentation* 367
Curry, C.: *Geometrical Optics* 364

Durand, E.: *Électrostatique et Magnétostatique* 591

Editions de la Revue d'Optique Theorique et Instrumentale 1953: *Colloque sur la Sensibilité des Crystaux et des Émulsions Photographiques, Paris*, 1951, *Mémoires et Discussions* 167
Editions de la Revue d'Optique Theorique et Instrumentale: *Comptes Rendus du Premier Congrès International de Microscopie Électronique* 653
Ettinger, G. M.: *Magnetic Amplifiers* 505

Finniston, H. M. (Ed.): *Metallurgy of the Rarer Metals* 795
Flügge, W.: *Four-Place Tables of Transcendental Functions* 653
Fowler, R. G.: *Introduction to Electric Theory* 509

Gomer, R., and Smith, C. S. (Ed.): *Structure and Properties of Solid Surfaces* . 851
Goodfriend, L. S. (Ed.): *Journal of the Audio Engineering Society* 365

Hall, E. O.: *Twinning and Diffusionless Transformation in Metals* 733
Hall, J. A.: *Practical Thermometry.* (Institute of Physics Monographs for Students) 274
Hallimond, A. F.: *Manual of the Polarizing Microscope* 508

Institute of Metals: *Properties of Metallic Surfaces* 502
Institute of Physics: *Static Electrification: A Symposium held by the Institute of Physics in London, 25th–27th March* 1953. (British Journal of Applied Physics Supplement No. 2) 594
Ivanov, A.: *Les aberrations de l'œil—leur rôle dans l'accommodation* 900

PAGE

Klemperer, O. : *Electron Optics* (2nd edn) 273
Kolsky, H. : *Stress Waves in Solids* 444

Lipson, H., and Cochran, W. : *The Determination of Crystal Structures* . . 655

Macfadyen, K. A. : *Small Transformers and Inductors* 445
McLachlan, N. W. : *Complex Variable Theory and Transform Calculus* . . 591
Meek, J. M., and Craggs, J. D. : *Electrical Breakdown of Gases* 594
Middleton, W. E. K. : *Vision Through the Atmosphere* 166
Miller, G. L. : *Metallurgy of the Rarer Metals.* No. 2, *Zirconium* . . . 795
Morgan, Joseph : *Introduction to Geometrical and Physical Optics* . . . 275
Muchmore, R. B. : *Essentials of Microwaves* 507

Newell, Homer E. : *High Altitude Rocket Research* 733

Pinsker, Z. G. : *Electron Diffraction* 278
Preston, R. D. : *The Molecular Architecture of Plant Cell Walls* 507

Radley, J. A., and Grant, J. : *Fluorescence Analysis in Ultra-Violet Light* . . 850
Ramo, S., and Whinnery, J. R. : *Fields and Waves in Modern Radio* . . . 854
Randall, J. T. (Ed.) : *Nature and Structure of Collagen* 732
Richards, C. E., and Lynch, A. C. : *Soft Magnetic Materials for Telecommunications* 654
Rogers, J. S. : *Physics for Medical Students* 799

van Santen, G. W. : *Introduction to a Study of Mechanical Vibration* . . . 504
Scott Blair, G. W. (Ed.) : *Foodstuffs, their Plasticity, Fluidity and Consistency* . 506
Smith, C. J. : *The General Properties of Matter* 86
Starling, S. G. : *Electricity and Magnetism for Degree Students* 798
Stiles, W. S. : *Physiology of Vision* (Vol. 9, No. 1, 1953, of British Medical Bulletin) 85
Sully, A. H. : *Metallurgy of the Rarer Metals.* No. 1, *Chromium* . . . 795
Sutton, O. G. : *Mathematics in Action* 734
Swainger, K. : *Analysis of Deformation.* Vol. 1, *Mathematical Theory* . . 596
Symon, K. R. : *Mechanics* 511

Terman, T. E., and Pettit, J. M. : *Electronic Measurements* 272
Thomas, T. S. E. : *Physical Formulae* 364
Trumpler, R. J., and Weaver, H. F. : *Statistical Astronomy* 902

United States Department of Commerce, National Bureau of Standards : *Tables of* 10^x (*Antilogarithms to the base* 10). (Applied Mathematics Series No. 27) . 510

Verma, Ajit Ram : *Crystal Growth and Dislocations* 446

Weinstock, Robert : *Calculus of Variations with Applications to Physics and Engineering* 84
Westphal, W. H. : *Physikalisches Wörterbuch* 510
Williams, A. O. : *Electronics : Physical Principles and Applications* . . . 734
Wilson, A. J. C. (Ed.) : *Structure Reports for* 1945–1946, Vol. 10 594
Woodward, P. M. : *Probability and Information Theory, with applications to Radar* 501
Woolley, R. v. d. R., and Stibbs, D. W. N. : *The Outer Layers of a Star* . . 446

Yarwood, T. M. : *Acoustics : a Textbook for Physics and Engineering Students* . 855